Correlating social mobility and economic outcomes

Citation for published version:
Guell, M, Pellizzari, M, Pica, G & Rodríguez Mora, JV 2018, 'Correlating social mobility and economic outcomes', *The Economic Journal*, vol. 128, no. 612, pp. F353-F403. https://doi.org/10.1111/ecoj.12599

Digital Object Identifier (DOI):
10.1111/ecoj.12599

Correlating Social Mobility and Economic Outcomes*

Maia Güell[†] Michele Pellizzari[‡] Giovanni Pica[§]
José V. Rodríguez Mora[¶]

First version: February 2015
This version: July 2017

Abstract

We construct comparable measures of intergenerational mobility (IM) for 103 Italian provinces using the methodology of Güell, Rodríguez Mora and Telmer (2007; 2014) and explore their correlation with a variety of social and economic outcomes. We find that higher IM is positively associated with economic activity, education and social capital, and negatively correlated with inequality. Moreover, there is no clear pattern of correlation with other socio-political variables. These results are qualitatively similar to Chetty, Hendren, Kline, and Saez (2014), with the important difference that Italy is a highly centralised state where institutions and policies are 'de jure' the same in all provinces. This suggests that something beyond institutional and policy differences also shapes intergenerational mobility.

Key words: Surnames, intergenerational mobility, cross-sectional data analysis.
JEL codes: C31, E24, R10

*We thank Cristina Blanco, Nicola Solinas, Alessia de Stefani and Robert Zymeck for superb research assistance. We thank Giacomo Giusti of the *Istituto Guglielmo Tagliacarne* for providing us with the province-level data on value added in 1981. We are also very grateful to Daniele Checchi, Carlo Fiorio and Marco Leonardi for sharing their program files to generate the traditional measures of mobility using the Bank of Italy SHIW data. We benefited from the very useful comments from two anonymous referees and the Editor, and seminar participants at Stanford University, Universidad de Alicante, University of Geneva, University of Manchester, EUI, Tinbergen Institute, the IZA-CEPR and EALE meetings, and the Madrid Mobility Workshop 2016. We are grateful to Joan Gieseke for editorial assistance. Financial support from the Spanish Ministry of Education and Science under grants ECO2011-28965 (MG) and ECO2011-25272 (JVRM) is gratefully acknowledged. Maia Güell is also grateful to the financial support from the Spanish Ministerio de Economia y Competitividad (MINECO-ECO2014-59225-P), and Michele Pellizzari gratefully acknowledges the financial support of NCCR-LIVES.

[†]University of Edinburgh, CEPR, FEDEA and IZA. Email: maia.guell@gmail.com
[‡]University of Geneva, CEPR, LIVES and IZA. Email: michele.pellizzari@unige.ch
[§]Università degli Studi di Milano, LdA, CSEF and Centro Baffi. Email: giovanni.pica@unimi.it
[¶]University of Edinburgh and CEPR. Email: sevimora@gmail.com

1 Introduction

Recent literature collects measures of intergenerational mobility (IM, hereafter) across different areas and correlates them with economic and social outcomes. Corak (2013b), for instance, compares IM across countries and documents that higher social mobility is associated with less inequality.[1,2] Chetty, Hendren, Kline, and Saez (2014) compare social mobility measures across U.S. commuting zones and find that higher mobility is associated with less segregation, less inequality, better schools, greater social capital and family stability. This evidence, while not causal, suggests that policy and institutional differences may be one of the underlying drivers of these correlations.[3] Clark (2014), instead, looking at the correlation across generations of the average outcomes of individuals sharing the same surname, claims that mobility does not vary much across societies, and it is therefore uncorrelated with economic conditions.

This paper contributes to this debate by looking at the correlation between IM and a variety of interesting social and economic outcomes across different geographical areas (provinces) of a single non-federal country, Italy, in which all provinces share the same institutional framework.[4] We measure IM by applying a novel method based on surnames proposed by Güell, Rodríguez Mora and Telmer (2007; 2014) to the universe of all tax declarations submitted in Italy in 2005. While Italy is one of the most immobile countries according to Corak (2013a), our estimates show that Italian provinces exhibit a large degree of variability in social mobility. We exploit this variability to explore the correlation between IM and an array of aggregate economic and social indicators and find that IM is higher in provinces where the level of economic activity is higher, inequality is lower, and social capital and educational attainments are higher. We also find that IM does not correlate in any systematic way with other socio-political variables, such as crime rates and life expectancy.

[1]See also Corak (2006) for an earlier analysis of a cross-country comparison between IM and the return to tertiary education, an important determinant of cross-sectional inequality.

[2]Less recently, Björklund and Jäntti (1997), Couch and Dunn (1997), Checchi, Ichino, and Rustichini (1999), Björklund, Eriksson, Jäntti, Raaum, and Österbacka (2002), Comi (2003) and Grawe (2004) compare mobility patterns across countries. In the literature review by Black and Devereux (2011) the authors offer a discussion of why IM might differ across countries.

[3]Björklund and Salvanes (2011) offer a review of recent empirical research on education and family background, which includes a discussion on the impact of educational policy on IM.

[4]Recently, Aydemir and Yazici (2015) provide correlates of IM and socio-economic development within Turkey and Heidrich (2015) studies the regional mobility patterns in Sweden.

Our work contributes to the literature in a number of dimensions. First, we confirm the evidence by Corak (2013b) and Chetty, Hendren, Kline, and Saez (2014) on a different country and using completely different data and methodology. Second, the fact that we exploit *within*-country variation and focus on Italy, a highly centralised country in terms of political institutions and policy making, allows us to conclude that the correlations that we document can hardly be explained by differences in policies, such as those related to education or welfare. This is an important contribution that differentiates our work from most papers in this area, such as Corak (2013b) and Chetty, Hendren, Kline, and Saez (2014), which compare political entities implementing very different policies that are likely to directly affect both the degree of IM and socio-economic outcomes. In Italy such policies are *de jure* decided at the central level. Hence, the correlations between IM and the vast array of outcomes that we consider cannot be attributed to differences in policies across Italian provinces. Something else, beyond local policies, must be jointly driving the degree of intergenerational mobility and macroeconomic outcomes.

From a methodological point of view, we follow Güell, Rodríguez Mora and Telmer (2007; 2014), who measure mobility by using an indicator – the *Informational Content of Surnames* (ICS) – of how much individual surnames explain the total variance of individual outcomes. The ICS compares the within-surname variance of an individual outcome, income in our case, with its unconditional variance. The lower the within-surname variance with respect to the unconditional one, the lower the degree of social mobility. This method allows us to construct IM measures for small geographical areas without relying on panel data. It measures mobility by looking at the imprint of inheritance on a cross section of individual outcomes. An intrinsically dynamic characteristic (mobility), which cannot be observed directly, can be inferred from its effect on an observable variable, namely the ratio of the within-surname to the unconditional variance of earnings.[5]

[5]It is important to stress that this methodology differs in many respects from other recent work using surnames to measure mobility, such as Collado, Ortuño-Ortín, and Romeo (2012), Clark (2014) or Diaz Vidal (2014). These papers use a "group estimator" of the standard IM coefficient, which in this context is likely to produce an upward bias (see Güell, Rodríguez Mora and Telmer (2007; 2014) for further discussion). More concretely, Clark (2014) averages individual outcomes within surnames for each generation – thus eliminating the within-surname variance of individual outcomes – and then looks at the correlation of those within-surname averages across generations. Clark (2014)'s procedure mechanically makes the unconditional variance of individual outcomes smaller and induces an upward bias in the estimate of how much surnames can explain such total variance. This is why he gets very high persistence rates in all countries. This shortcoming of the group

Güell, Rodríguez Mora and Telmer (2007; 2014) show theoretically that the ICS maps into the standard measures of mobility. Moreover, they show empirically that the evolution of IM over time in Spain mimics the evolution of standard sibling correlations. In this paper, we further corroborate the association between the ICS and IM by showing the similarity in our findings with Corak (2013b) and Chetty, Hendren, Kline, and Saez (2014), who use administrative data.

The paper is organised as follows. Section 2 describes the methodology based on the informational content of surnames used to measure intergenerational mobility across Italian provinces. Section 3 provides information on the rules governing the transmission of surnames in Italy. Section 4 describes the data used; Sections 5 and 6 discuss the results of the analysis. Section 7 concludes.

2 Measuring Mobility

In this paper, we use the measure of intergenerational mobility proposed by Güell, Rodríguez Mora and Telmer (2007; 2014), the *Informational Content of Surnames* (ICS). Unlike traditional measures of mobility, it does not require panel data nor any explicit links between children and their parents. One cross-sectional data set of surnames and economic outcomes is enough.[6]

Our approach has many similarities to the well-established methodology of looking at sibling correlations in order to infer IM. If economic inheritance is important, the outcomes of siblings should be correlated because they share parents and, thus, they share the same inherited economic traits. Consequently, the variance of siblings' income should be similar to the

estimators is well-known and has been shown empirically by Chetty, Hendren, Kline, and Saez (2014), and explained by Solon (2016) and Vosters (2017). The approach we follow does not suffer from this bias because it uses individual-level (as opposed to surname-level) outcomes. In Appendix A we estimate mobility using Clark's method on the Italian data of this paper and, as Chetty, Hendren, Kline, and Saez (2014), we show that a large correlation of (within-surname) income averages across generations can be obtained only by focusing on very common surnames, that capture geographical income differences rather than social mobility.

[6]The ICS also differ substantially from the so-called *Two-Sample Two-Stages Least Squares* (TS2SLS), an alternative methodology used by some authors to overcome the need of long panels to compute empirical measures of IM (Björklund and Jäntti, 1997; Barone and Mocetti, 2016). The ICS only requires one simple cross-section of data whereas TS2SLS requires representative data on at least two generations. In particular notice that having data on two cross-sections representative of the same population at two different points in time might not necessarily provide representative data on two generations or birth cohorts. Hence, we see the data requirements of the TS2SLS as being substantially stronger than those of the ICS.

population variance if inheritance is irrelevant, but much smaller if inheritance matters a lot. If income follows an AR(1) process with autocorrelation ρ and conditional variance σ^2, the ratio of sibling variance to total variance is

$$\frac{\sigma^2}{\frac{\sigma^2}{1-\rho^2}} = 1 - \rho^2.$$

Notice that this ratio is the R^2 of a regression of individual income on sibling dummies. This works in an obvious manner for siblings, because we know the exact relationship between them and with the ancestor from whom they get inheritance. Essentially the same procedure works using surnames because surnames establish a partition of the population that is informative about family links.

The amount of information contained in surnames is the ratio of the variance of income conditional on sharing a surname to the unconditional variance of income - that is, the R^2 of a regression of individual income on surname dummies. Given a certain mapping between the surname partition and family linkages, the more prevalent inheritance is, the larger the amount of information that surnames will contain.

Thus, the key to the method is that surnames are informative about family linkages. They do happen to be informative because surname distributions are very skewed. If there were only a few surnames, the mapping between the surname partition and family relationship would be extremely blurred, and conditioning on surnames would not change the variance for any degree of inheritance. Fortunately, Western surname conventions ensure that surname distributions are bound to be very skewed. Despite the presence of a small number of surnames shared by very many people – who are very unlikely to have common ancestors – surname distributions typically contain a very large number of uncommon surnames shared by few individuals who are instead very likely to have close family relationships. In those infrequent surnames lies the power of the methodology.

2.1 The Informational Content of Surnames

The *Informational Content of Surnames* is a measure of how much information surnames contain about the economic outcomes of individuals, after controlling for other factors. In this section we describe the ICS in detail.

Consider a cross section in which each individual is associated with a surname s, a measure of economic well-being y_{is}, and a vector of additional demographic characteristics X_{is}, such as age and gender. Güell, Rodríguez Mora and Telmer (2007; 2014) define the ICS as the difference between the R^2 of two regressions. The first regression, whose R^2 is denoted as R^2_L, models the economic well-being of individual i with surname s as follows:

$$y_{is} = \gamma' X_{is} + b' D + \text{residual}, \tag{1}$$

where D is an S-vector of surname-dummy variables with $D_s = 1$ if individual i has surname s and $D_s = 0$ otherwise.

Since the number of surnames is very large and they may happen to explain the variance of y_{is} even if they do not carry any information on family linkages, a second set of regressions is performed to ensure that we do not spuriously attribute informativeness to surnames. In each of the regressions, we include a different S-vector of 'fake' dummy variables F that randomly reassign surnames to individuals in a manner that maintains the marginal distribution of surnames but destroys the informativeness of surnames about familial linkages. The regression is

$$y_{is} \quad = \quad \gamma' X_{is} + b' F + \text{residual}. \tag{2}$$

The R^2 from this regression is denoted as R^2_F. We replicate the regression in (2) ten times and calculate the average of all the R^2 obtained.[7] Denoting such an average as $\overline{R}^2_F$, the ICS is defined as

$$\text{ICS} \equiv R^2_L - \overline{R}^2_F. \tag{3}$$

[7]Results do not depend on the number of replications.

The ICS measure has a number of important properties. It has value zero if there is one surname per person or if there is only one surname for everyone. More generally, it captures the information that surnames contain because of family linkages and measures how much of the variance of the dependent variable is explained by the variance of the surnames.[8]

2.2 Cross-provincial comparability of the ICS

Given that our goal is to get comparable estimates of the ICS for each Italian province in order to investigate the correlation between mobility and a battery of aggregate socio-economic outcomes, it is of paramount importance that the distributions of surnames across provinces are comparable so that any differences in the ICS reflect differences in social mobility and not in other factors.

Section 5 shows that the distributions of surnames are indeed very similar across provinces once we drop individuals with surnames that are too frequent to be informative about family connections. The tail of the surname distribution that contains infrequent surnames identifies family linkages with less noise and is therefore more comparable across provinces. For this reason, in the paper, we will use the ICS computed on individuals whose surname contains less than 30 people in the province as a baseline measure of social mobility and show that results are similar both when using all individuals and when concentrating on individuals whose surname contains less than 15, 20 and 25 people. This issue is further discussed in Section 5.

An additional challenge that may affect the cross-province comparability of the ICS is migration, both from other countries and from other Italian provinces. Migrants may have both very different surnames and very different economic outcomes as compared with natives in the recipient region (at least initially). Hence, their surnames might be very informative regardless of the degree of IM in the province. Additionally, if highly motivated young people in southern Italy move to the North or emigrate, this may raise within-family income correlation in the South with respect to the North. Unfortunately, since our data do not include information on the birthplace of the individuals, we cannot directly track migrants.

[8] Güell, Rodríguez Mora and Telmer (2007; 2014) provide a model that maps the ICS into the traditional measure of IM based on father-son regressions and show that the former is monotonically increasing in the latter.

Following Güell, Rodríguez Mora and Telmer (2007; 2014), we can, however, construct an index of the local dimension of surnames and focus our analysis on the individuals whose surname is relatively common in their province of residence. Such individuals are very unlikely to be migrants. We measure how local a surname s in province r is as follows:

$$LocalDegree(s,r) = \frac{\text{Number of people with surname } s \text{ in province } r}{\text{Number of people with surname } s \text{ in Italy}} \tag{4}$$

To the extent that migrants have very different surnames from natives, they display a low value of the index in the recipient province. Therefore, by restricting the analysis to individuals whose surnames are local enough, we plausibly exclude immigrants and minimise the effect of migration in the province of *destination* on the ICS.

Yet, this procedure does not resolve the migration issue completely because it allows us to identify likely migrants in the province of destination (and drop them) but is silent about their origin. For this reason, our estimates of social mobility in the provinces from which individuals migrate may still suffer a bias because we do not observe the individuals that left. To account for this potential bias when looking at the correlation between mobility and macroeconomic outcomes, we perform a number of robustness checks. First, given that internal migration in Italy mostly flows from the South to the North, we include a North/South dummy in the regressions of IM on province-level outcomes. Second, we control for the net province-level migration flows obtained from the Italian National Institute of Statistics. None of these robustness checks change our results significantly.

3 Italian Surnames

In Italy, surnames follow the standard Western naming convention. Most people inherit their surnames from their fathers. At the same time, there can be some surname innovations because it is possible, although not easy, to change one's surname. The procedure to do so is quite complex and can take up to one year. As discussed in Güell, Rodríguez Mora and Telmer (2007; 2014), this naming convention implies that the resulting distribution of surnames is very skewed, meaning that most people have very infrequent surnames and that the likelihood that any two persons sharing an unusual family name are linked by some family connection is extremely high.

Unlike most other countries, in Italy women do not change their official surnames upon marriage. While in everyday life it may happen that married women use their husbands' surname, the law requires everyone to use their inherited surnames in all official documents regardless of marital status. Indeed, in Italy the government identifies taxpayers through a unique fiscal code (*codice fiscale*), which is given to each person at birth and does not change with marriage. The code depends on the name, the surname at birth, date and place of birth. So, the state identifies taxpayers through the surname at birth. Furthermore, the instructions for income tax forms state explicitly that married women should use their maiden surname.

As already mentioned, it is possible to change one's surname, in which case one's fiscal code is also changed. This same procedure also applies to married women who want to officially add their husbands' surnames to their original ones or even replace their maiden surnames with their husbands'. Hence, in the vast majority of cases, both men and women file their tax reports using their inherited surnames. This means that technically we can calculate the ICS for the entire population, both males and females, using tax data. In our baseline estimates, we focus on males, as most of the literature does. Appendix F provides estimates that include females as well.

4 Data

In this paper, we exploit very rich individual-level microdata from Italy with information on both individual surnames (anonymised) and individual taxable incomes. We use these data to compute measures of the ICS at the provincial level. We then link such measures with macroeconomic variables at the same level of geographical aggregation.[9] We obtain these macroeconomic variables from a variety of different sources.

[9] The exact number and boundaries of the provinces have changed a few times over the recent decades. We use the definition of provinces as of 2004, which is the reference year of our tax data, although the current (2016) definitions are slightly different.

4.1 Tax records

Our main indicators of mobility are the ICS computed by using data from the universe of all the official tax declarations in Italy for the year 2005. These declarations were submitted between the beginning of May and mid-June 2005 and refer to all taxable incomes (excluding capital incomes) earned between January 1 and December 31, 2004. We obtained the data from the website of the Italian Ministry of Finance, where they were published on April 30 2008, but were subsequently removed following the intervention of the Italian Privacy Authority. Even though the individual tax declarations were (and still are) classified as public information in Italy, the procedure to access them is strictly regulated and the Authority deemed that the online publication did not conform to the law. The formal procedure to access the data requires submitting an individual request to the local branch of the tax authority, which can provide information exclusively regarding the citizens who reside in its area.

The Authority also clarified that whoever had obtained the data through the Ministry's website had done so legally. However, the norms regulating access to the data apply to everyone and it is prohibited to distribute them, at least in their original format, other than through the formal legal procedure. For this project, we have produced a fully anonymised version of the data, with individual names and surnames replaced by numerical codes (still allowing for the identification of individuals sharing the same names or surnames), which we use to produce all of the results in the paper and which can be distributed for replication purposes. The same data have been used by Braga, Paccagnella, and Pellizzari (2016); Anelli and Peri (2013). The very special situation under which the 2005 tax records were made available did not reproduce itself, and only this year of data is available for research purposes. Researchers at some institutions, such as the Ministry of Finance or the Bank of Italy, might have access to more detailed data covering longer time periods under special agreements (see, for example, Barone and Mocetti (2016); Mocetti and Viviano (2015)).

Despite covering the entire universe of submitted declarations, our data do not necessarily include the whole Italian population. Although in principle every resident in Italy is required to submit a tax declaration, there are exceptions. The first and most important exception includes children (and any other dependent family members), who are not required to submit their own tax forms but appear in the forms of their parents (either one or both) who may be eligible for family allowances.[10] The second important category includes persons whose income falls below a given threshold, who are exempted from declaring taxes. The exact threshold depends on the composition of the income sources and varies between €3,000 and €7,500 in the year of our data. Among this second group of exemptions are also those who earn exclusively capital income, which is taxed separately in Italy and does not enter the calculation of personal taxable income.

Italy has three different forms of tax declarations. Persons who only have incomes from dependent employment have their taxes deducted directly from their monthly salaries, and their employers submit a summary tax report for them. Technically, these persons do not submit any form themselves. The second form is used by those who have incomes from both dependent employment and other sources. Finally, the third form is for all those who do not fall into either of the first two groups, namely the self-employed and those with incomes from rents and dividends. In our data, each of these forms is used by about one-third of the taxpayers.

All three tax forms are quite voluminous, from 6 to 30 pages depending on the exact situation of the taxpayer. However, our data contain only a limited subset of this information, namely the names of the person submitting the file, their dates of birth, the province of residence, total taxable income, the most prevalent source of income (e.g., dependent employment, self-employment, rents and dividends), the amount of the tax due and the form used for the declaration.

In the original data, the first name and the surname of the taxpayer are coded in a single string variable, and in order to separate them, we used the following procedure. First, we considered only those cases in which the original string contained only two separate words, indicating that the person only has one name and one surname. For these cases, we know that the first word is the first name and the second is the surname. About 70% of cases in our data were settled in this simple way. For the others, we created an archive of first names using those derived in the first step of our procedure, complemented by a number of additional

[10] Technically, one is considered a dependent family member if one's income is below a fixed threshold (€2,840.51 in 2004). Submitting one's own declaration separate from that of the household head is, however, always possible.

lists of Italian first names.[11] Next, we considered records with more than two words in the original string variable, and we coded as surnames the continuous sequences of words that did not appear in our archive of first names. The sequences are continuous in the sense that the algorithm takes into account the fact that the original string must be formed by a sequence of first names followed by a sequence of surnames and the two cannot be mixed. We then coded the remaining sequences of words as first names. Our archive of first names also allowed us to classify them by gender, although about 7.5% of the records could not be unambiguously assigned to a gender.[12]

Overall, there are 38,514,292 records in the original tax files, which compares with about 50 million residents in Italy aged 16 and over in 2004 or about 80% of the entire population who could legally earn labour incomes.[13] In order to limit complications arising from the process of labour market participation, we focus exclusively on men and we drop observations for which the information on gender is not reliable. This leads to approximately halving the original population. Further, we exclude outliers aged above 100 years and individuals with unique surnames in their province, for whom the ICS is not defined. This leaves us with 18,890,891 observations, of which 18,884,811 have nonmissing taxable income.

Taxable income, as recorded in the tax declarations, is our main indicator of economic success and the basis for our analysis of mobility. According to Italian legislation as of 2005, taxable income is the sum of all gross earned incomes (excluding capital income) minus deductions, which are granted for a variety of reasons (e.g., number of children, mortgage interest on first homes, some medical and educational expenses, and so on). Importantly, the rules defining fiscal deductions do not vary across geographical areas. These allowances, plus the fact that the self-employed can report losses, mean that taxable income can be zero. The existence of the allowances also implies that individuals with the same taxable income may end up paying different amounts of taxes.[14]

[11]For this, we use the first names of lawyers and politicians (who are all registered in public registries where first names and surnames are clearly separated) and a number of websites and books providing guidance to parents who are choosing a name for their newborns.

[12]These ambiguities are much more likely to arise for foreigners than for Italians.

[13]Education in Italy is compulsory until the age of 15, so 16 is the minimum working age.

[14]In Güell et al. (2015) we present results based on ICSs computed using the net tax paid instead of taxable income as an indicator of economic success, and results are unchanged.

Table 1. Tax records: descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
	Panel A: individual-level			
Taxable income	15,737.21	42,993.09	0	101,255,692
	Panel B: surname/province-level			
Number of individuals in the province per surname (a)	16.32	60.43	2	18,684
Number of individuals in the province (b)	334,004.3	35,3625.6	30,632	1,249,617
Frequency of surname (a/b) ($\times$ 10,000)	0.890	2.815	0.016	237.199

The statistics in Panel A are computed on 18,884,811 individual observations. The statistics in Panel B are computed on 1,157,740 surname-province observations.
Source: 2005 Italian tax records. Population: men aged 16-100 years old.

Table 1 (Panel A) reports some descriptive statistics for our data. The final working population contains about 19 million taxpayers with an average annual gross income of about 15,500 euros and a standard deviation of almost 43,000 euros, approximately 2.8 times the average. A nonnegligible fraction of individuals, around 18% in our population, declare zero income. Given the size of this group, we want to keep in our the empirical analysis; hence, we take the log of (1+taxable income) as a dependent variable in regressions 1 and 2. As is common with most distributions of incomes, we see a relatively long right tail, with the 95th percentile at around 50,000 euros and the 99th percentile just over 100,000 euros.

Tax evasion is a well-known phenomenon of the Italian economy, and it is reasonable to think that these fiscal records are only noisy measures of the true underlying incomes (Fiorio and D'Amuri, 2005). In Section 4.1.1 we discuss this issue and its potential implications for our empirical exercise.

For the purpose of constructing the ICS, the distribution of surnames is perhaps more interesting than the distribution of individuals (Table 1, Panel B). We have about 1 million surnames (treating the same surname in different provinces as different units) with 16.32 individuals holding the same surname on average in the same province. Considering that the average province has about 334,000 residents, each surname covers on average slightly less than 1 (0.890) every 10,000 persons. This (very low) average frequency approximates the probability that any two individuals taken at random in a typical province share the same surname. Instead, the probability that two individuals at random belong to the same family conditional on having the same surname is, in the typical province where each surname contains 16.32 individuals, given by family size/16.32. Taking the extremely restrictive view that the average family size is equal to 3, this probability is equal to $3/16.32 = 0.1838$, that is about 2,000 times larger than the probability that any two individuals taken at random share the same surname and about about 20,000 times larger than the probability that any two individuals taken at random (unconditional on sharing surnames) belong to the same family. As predicted by the rules of surname transmission, the distribution of surnames is very skewed. The median frequency of surnames is 1 every 40,000 and the 25% percentile is 1 every 90,000.

4.1.1 Under-reporting

Given the large degree of tax evasion in Italy, it is reasonable to think that the incomes in the fiscal declarations are often under-reported.[15] A more severe form of tax evasion is not filing a tax declaration altogether but this is a very rare phenomenon even in Italy as it essentially implies major difficulties accessing a number of important public services, such as health care and education. In this section, we briefly discuss the role of under-reporting for the computation of the ICS assuming that all residents who are required to submit a declaration do so.

In our case, one might worry that under-reporting could affect our results depending on its pattern across income and geographical distributions. One can imagine, for instance, that richer provinces may have less or more under-reporting than poorer regions. Note, however, that any differences in the *level* of under-reporting across provinces do not influence the ICS, which measures the ratio between the conditional and unconditional *variances* of income. Thus, for under-reporting to affect the ICS, it needs to differentially affect the variance of income across provinces. If, for example, under-reporting generates noise and therefore raises the unconditional variance of income, in provinces in which incomes are more often under-reported, the measured ICS will be lower.

To see this argument more formally, let us assume that, because of under-reporting, the true income y_{isp} of individual i with surname s in province p appearing in equation (1) is not

[15]The Italian National Institute of Statistics (ISTAT) estimates that in 2004 – the year of the incomes used in this paper – the undeclared economy ranged between a minimum of 16.1% to a maximum of 18.1% of national GDP and that about 10.1% of employed workers were undeclared (i.e., their contracts were not registered and/or they were not paying social security contributions).

perfectly observable and we only observe a noisy version of it, namely $y^*_{isp} = \alpha_p y_{isp} + \varepsilon_{isp}$, with $\alpha_p < 1$ measuring the province-specific evaded proportion of income and ε_{isp} being an error term uncorrelated both with the true level of income of the individual and with his or her surname.[16]

The term α_p has clearly no impact on the ICS because it does not affect the province-specific R^2 of the estimates of equation (1): it is just a rescaling factor. The term ε_{isp} may have an impact on the ICS only if its variance is province-specific.[17] The reason is that in provinces in which the variance of ε_{isp} is larger the unconditional variance of income is also larger and the R^2 from the estimates of equation (1) necessarily lower. Instead, a province-specific expected value of ε_{isp} is absorbed by the province-specific constant of the regression and does not affect the ICS.

Thus, whether tax evasion affects our results is ultimately an empirical matter and depends on whether differential under-reporting across provinces affects the unconditional variance of the observed incomes. To address this issue, we exploit differences in the likelihood of under-reporting across individuals earning incomes from different sources. In fact, individuals who only earn income from dependent employment are taxed at the source by their employers and cannot choose to under-report. Hence, tax evasion is mostly an issue of the self-employed. Appendix F investigates empirically how the spatial distribution of the estimates of the ICS is affected by under-reporting excluding the self-employed – who are seemingly more prone to under-report – from the analysis and finds that all the correlations with the macroeconomic variables remain virtually unchanged.

Notice also that if (for whatever the reason) misreporting were more prevalent in the South, as some people may suggest, the ICS would be relatively underestimated in the South. Given that we find the opposite (the ICS is substantially higher in the South), this effect would mean that we are underestimating the differences.

[16]In order to highlight the role of the heterogeneity across provinces, here we add the subscript p, whereas it is omitted in equation (1) for brevity.

[17]This may happen even if ε_{isp} is uncorrelated to surnames and does not bias the estimates.

4.2 Macrodata

For each of the 103 provinces, we collect various aggregate economic and social outcomes. These data come from the Italian National Institute of Statistics (ISTAT), unless otherwise explicitly specified below. Our ICS indicators are produced using data on incomes earned in 2004.

Ideally, one would like to relate these data not only to recent economic outcomes but also to outcomes decades ago. Although this approach would not allow us to go beyond simple correlations between social mobility and macroeconomic outcomes, it would enhance our understanding of how persistent the correlations are, given that mobility is arguably a slow-moving variable. Unfortunately, ISTAT does not provide province-level variables for the years prior to 1999. For this reason, most of our variables refer to the period 1999-2004. As a notable exception, we have value added per capita in 1981 kindly made available by the *Istituto Guglielmo Tagliacarne*. Table B1 in Appendix B lists all the variables and specifies the years for which they are available. To limit the impact of cyclical fluctuations and concentrate on long-run structural correlations, we average these variables over all available years whenever possible.

For the sake of clarity of exposition, we organise all of our province-level variables into three categories. The first category (labelled "key outcomes") contains outcomes that are of particular interest for the debate on the causes and consequences of low social mobility, such as the level of economic activity, educational attainment, inequality and social capital. The second category (labelled "other economic outcomes") refers to economic variables measuring labour market outcomes and the degree of trade openness of the province. The third group of variables (labelled "other socio-political outcomes") includes variables such as life expectancy, suicide rates, crime rates and public sector activity. The latter consists of variables capturing the degree of intervention of both the central and the local governments (value of public works started and completed, by either the central or the local government) and the efficiency of local governments (delay of payments to suppliers, measured by the ratio between paid and committed outlays in the municipal budget within the year, schooling level of the local politicians and the budget deficit).

Tables 2, 3 and 4 provide descriptive statistics for each group of variables. Without going into the details of each variable, it is worth noticing the great deal of heterogeneity that

Table 2. Key outcomes: descriptive statistics

	mean	Percentiles 10	50	90
Economic activity				
Value added per capita (avg 1999-2004)	18,830	11,932	19,378	24,717
Value added per capita (1981)	3,997	2,569	4,233	5,123
Educational attainment				
Individuals aged 25-64 with at most 8 years of schooling per 100 same-age individuals	52.84	44.96	52.61	61.58
Early school dropout aged 18-24 per 100 same-age individuals	22.26	14.32	21.54	31.88
Inequality				
Standard deviation of log income	3.985	3.60	3.92	4.40
Social Capital				
Voter turnout in Chamber of Deputies elections per 100 voters	82.05	74.86	83.23	87.47
Voter turnout in Senate of the Republic elections per 100 voters	82.17	74.54	83.18	87.58
Voter turnout in European Parliament election per 100 voters	73.94	63.09	75.12	81.09
Newspaper sales per capita	0.234	0.0540	0.130	0.481

Notes: All variables are available for 103 provinces, except for value added per capita in 1981 which exists only for 95 provinces. Table B1 describes the sources and years available of each variable.

characterises the Italian provinces. For example, value added per capita is on average equal to €18,830 (Table 2). However, the province at the 90^{th} percentile (Brescia) is 30% above the average, namely €24,717, and the province at the 10^{th} percentile (Trapani) is 37% below, namely €11,930. Thus, value added per capita is twice as large in Brescia as in Trapani.[18] The large degree of heterogeneity also characterises the distribution of the other variables in Table 2, including social capital (such as voter turnout and newspaper sales), education and cross-sectional inequality, and in Tables 3 and 4, with the exception – perhaps not surprisingly – of life expectancy.[19]

5 Surname distributions and the ICS

We use the Italian tax records described in Section 4.1 to obtain the surname distributions of Italian taxpayers for each province. To our knowledge, this is the most complete data set with (anonymised) surnames available for Italy, the closest to a census. To the extent that those distributions – the complex result of fertility processes, (assortative) mating and migration

[18] The number of observations reflects the number of provinces at the time each variable is measured: 95 in 1981 and 103 in the period 1999-2003.

[19] Our data, of course, confirm the well-known fact that provinces in southern Italy perform worse than those in the centre and in the North in terms of economic outcomes. They also confirm that the North/South divide in terms of value added per capita – for which we have data both for 1981 and for the beginning of the 2000s – is persistent, with little or no convergence taking place across provinces.

Table 3. Other economic outcomes: descriptive statistics

	mean	Percentiles 10	50	90
Economic activity				
Protested cheques per 1,000 inhabitants	564.5	211.9	460	1,034
Labour market outcomes				
Unemployment rate	9.322	3.237	5.854	21.53
Unemployment rate - Males	6.725	1.933	3.921	16.39
Unemployment rate - Females	13.75	4.931	8.877	31.40
Unemployment rate (age 15-24)	25.95	8.715	18.20	54.33
Long-term unemployment rate (12 months or more)	3.850	0.962	2.136	9.238
Employment rate	45.22	34.92	47.40	52.70
Employment rate - Males	56.73	48.97	57.68	63.62
Employment rate - Females	34.52	21.75	37.40	42.42
Employment rate (age 15-24)	28.46	13.43	31.23	41.37
Employment rate (high school, age 25-64)	73.69	60.43	76.87	82.02
Employment rate (at least college graduate, age 25-64)	79.61	72.53	80.18	85.48
Participation rate (age 15-64)	61.24	52.03	63.37	68.57
Participation rate (age 15-64) - Males	73.82	69.61	74.11	77.44
Participation rate (age 15-64) - Females	48.64	33.30	51.31	59.75
Participation rate (age 15-24)	32.92	24.05	33.03	40.92
Trade openness				
Imports to value added	172.9	38.74	152.9	315.9
Exports to value added	204.0	35.46	194.9	412.9

Notes: All variables are available for 103 provinces. Table B1 describes the sources and years available of each variable.

Table 4. Other socio-political outcomes: descriptive statistics

	mean	Percentiles 10	50	90
Life Expectancy				
Life expectancy at birth - Males	77.45	76.27	77.53	78.60
Life expectancy at 65 - Males	17.05	16.37	17.07	17.70
Life expectancy at birth - Females	83.22	82.27	83.30	84.13
Life expectancy at 65 - Females	20.99	20.33	21.07	21.67
Suicide Rates				
Suicides per 100,000 inhabitants - Total	7.272	3.887	6.954	10.99
Suicides per 100,000 inhabitants - Males	10.19	2.361	9.788	17.14
Suicides per 100,000 inhabitants - Females	2.950	0.474	2.645	5.583
Suicide attempts per 100,000 inhabitants - Males	7.129	2.043	5.607	16.86
Suicide attempts per 100,000 inhabitants - Total	7.621	3.213	6.393	13.58
Suicide attempts per 100,000 inhabitants - Females	7.401	1.211	5.163	18.55
Crime Rates				
Total crimes	3,520	2,409	3,284	5,106
Violent crimes	162.1	110.4	146.6	219.8
Thefts	1,932	1,013	1,775	3,106
Other crimes	1,467	1,040	1,410	1,948
Murders per 100,000 inhabitants	1.217	0	0.919	2.439
Petty thefts per 100,000 inhabitants	163.1	21.03	105.7	368.9
Snatching per 100,000 inhabitants	27.03	4.798	15.65	62.87
Burglaries per 100,000 inhabitants	425.0	225.1	398.4	588.2
Theft of parked cars per 100,000 inhabitants	355.5	152.2	304.4	622.5
Car thefts per 100,000 inhabitants	231.7	68.66	149.1	496.0
Scams per 100,000 inhabitants	123.8	73.07	117.3	168.9
Smuggling offences per 100,000 inhabitants	12.54	0.319	1.114	28.57
Drug production and sale per 100,000 inhabitants	63.07	27.98	52.59	97.00
Exploitation of prostitution per 100,000 inhabitants	4.767	1.729	3.611	8.146
Distraints per 1,000 inhabitants aged 18 years and older	8.026	3.434	7.238	13.47
Distraints per 1,000 families	17.06	7.393	15.12	27.59
Public Sector Activity				
Value of public works started (pct of VA)	17.36	4.517	10.23	25.22
Value of public works started by Provincial institutions (pct of VA)	0.867	0	0.267	1.764
Value of public works started in the construction sector (pct of VA)	3.113	1.042	2.477	5.525
Value of public works completed (pct of VA)	12.39	5.151	9.825	20.30
Value of public works completed by Provincial institutions (pct of VA)	0.644	0	0.295	1.631
Percentage politicians with at least secondary education	0.0232	0.0200	0.0230	0.0271
Ratio of paid to committed expenses	77.58	73.89	77.82	80.49
Deficit per capita in euros	12.17	3.889	11.66	22.82
Growth rate of deficit per capita in euros ($\times$100)	-5.030	-108.1	-0.717	14.05

Notes: All variables are available for 103 provinces, except for the Ratio of paid to committed expenses which exists for 102 provinces. Table B1 describes the sources and years available of each variable.

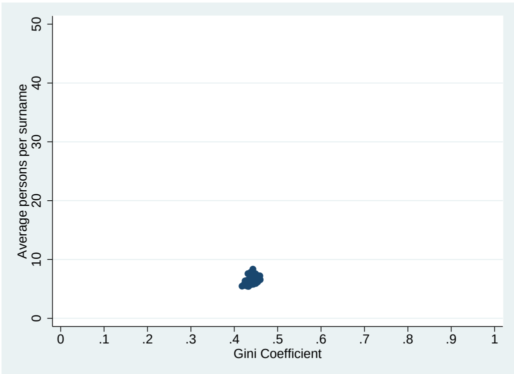
patterns – are similar, any differences in the ICS reflect differences in intergenerational mobility.

To investigate the similarity of the surname distributions across provinces, we exploit the well-known result that such distributions can be approximated very precisely by the Pareto distribution, which is uniquely characterised by two simple moments, the Gini coefficient and the number of persons per surname (Fox and Lasker (1983)). In other words, each pair of Gini coefficient and number of persons per surname uniquely identifies one surname distribution. We then calculate these two moments for each province and plot them in Panel (a) of Figure 1. If surnames were distributed identically in all provinces, the dots in the figure would overlap perfectly. This is clearly not the case in our data. While the Gini indices seem relatively homogeneous within the range $[0.6, 0.9]$, the average number of persons per surname spans between 10 and 50.

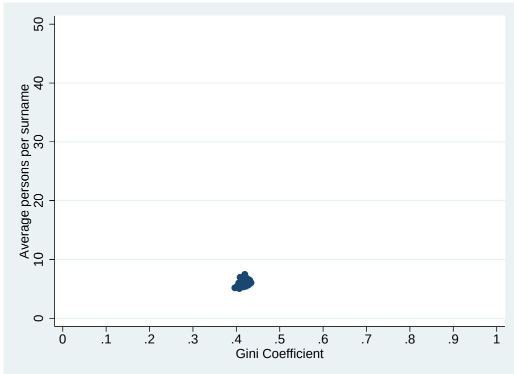
To enhance cross-province comparability, we then concentrate on the right tail of the distribution of surnames; that is, we focus on the individuals whose surnames are shared by less than a certain number of people (we experiment with 30, 25, 20 and 15). The idea behind this strategy is that, for these sub-populations, surnames measure family linkages more precisely. Panels 1(b) to 1(e) in Figure 1 show the Gini coefficient and the number of persons per surname for various tails of the distribution. These figures show that, once the most frequent, and thus the least informative, surnames are dropped, the surname distributions are virtually identical across all Italian provinces.

Based on this evidence, we are quite confident that, when using the tails, surnames map family relationships in similar manner in all provinces and that the mapping from the ICS to income persistence is thus comparable across provinces. For this reason, our baseline ICS measure is based on individuals whose surnames are shared by less than 30 people in their province. We label this indicator ICS-30. Results are robust to this choice. In Appendix E, we also show results using the full ICS – calculated on the entire distribution of surnames – and using the *Local* ICS-30, that is, the ICS computed on individuals whose surname contains less than 30 people and who belong to the 50% of the population with the most local surnames to partially account for differences in migration patterns across provinces (see Section 2.2).

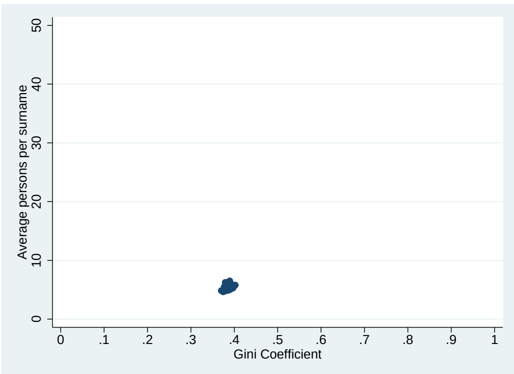
(a) All Individuals



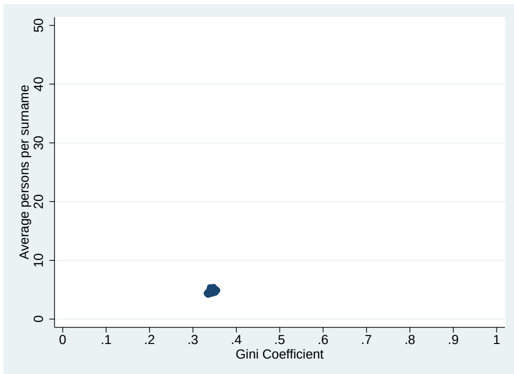
(b) Individuals with surnames with < 30 people



(c) Individuals with surnames with < 25 people



(d) Individuals with surnames with < 20 people



(e) Individuals with surnames with < 15 people

FIGURE 1. Comparability of surname distributions across provinces.

Table 5. ICS measures based on taxable income: descriptive statistics

	N	Mean	St.Dev.	Percentiles 10	50	90
ICS based on taxable income,	103	0.0247	0.0087	0.0151	0.0236	0.0370
ICS based on taxable income, tail 30 (baseline)	103	0.0456	0.0171	0.0289	0.0389	0.0724
ICS based on taxable income, tail 25	103	0.0478	0.0179	0.0311	0.0406	0.0751
ICS based on taxable income, tail 20	103	0.0505	0.0190	0.0332	0.0426	0.0802
ICS based on taxable income, tail 15	103	0.0540	0.0205	0.0351	0.0456	0.0842

Source: 2005 Italian tax records. Population: males aged 16-100 years old.

5.1 Empirical measures of the ICS

This section presents the empirical estimates of the mobility measures described in Section 2. Descriptive statistics for ICS measures based on taxable income are reported in Table 5. The first row refers to the ICS calculated on the full population, and the other rows report the ICS restricting the population to the individuals with the least frequent surnames (i.e. those containing less than 30, 25, 20, and 15 persons). Overall, the table shows that there is substantial variation in the ICS across provinces: the ICS-30 (our baseline measure) of the province at the 90^{th} percentile (Udine) is 2.5 times higher than the ICS of the province at the 10^{th} percentile (Agrigento). Not surprisingly, the level of the ICS monotonically increases when focusing on more and more infrequent surnames, because these are the ones that are the most informative about family linkages.[20]

Figure 2 provides a geographical breakdown of the estimates and shows that mobility increases when moving from the South towards the North of the country. Identifying the exogenous drivers of this geographical pattern is beyond the scope of this paper. Instead, we exploit the large geographical heterogeneity across Italian provinces to study how social mobility correlates with a number of macroeconomic outcomes (in Section 6), without necessarily making causal claims.

Table 6 shows descriptive statistics for ICS measures calculated for the fraction of individuals in the top 50% of the distribution of the $LocalDegree(s, r)$ Index in every province. As discussed in Section 2.2, this approach allows us to (partially) purge the ICS from the effect of migration in the provinces of destination.[21] From the second row on, we further restrict the population

[20]The underlying data are shown in Table C1.

[21]We will further address this problem in Section 6 including a North/South dummy in the regressions of the

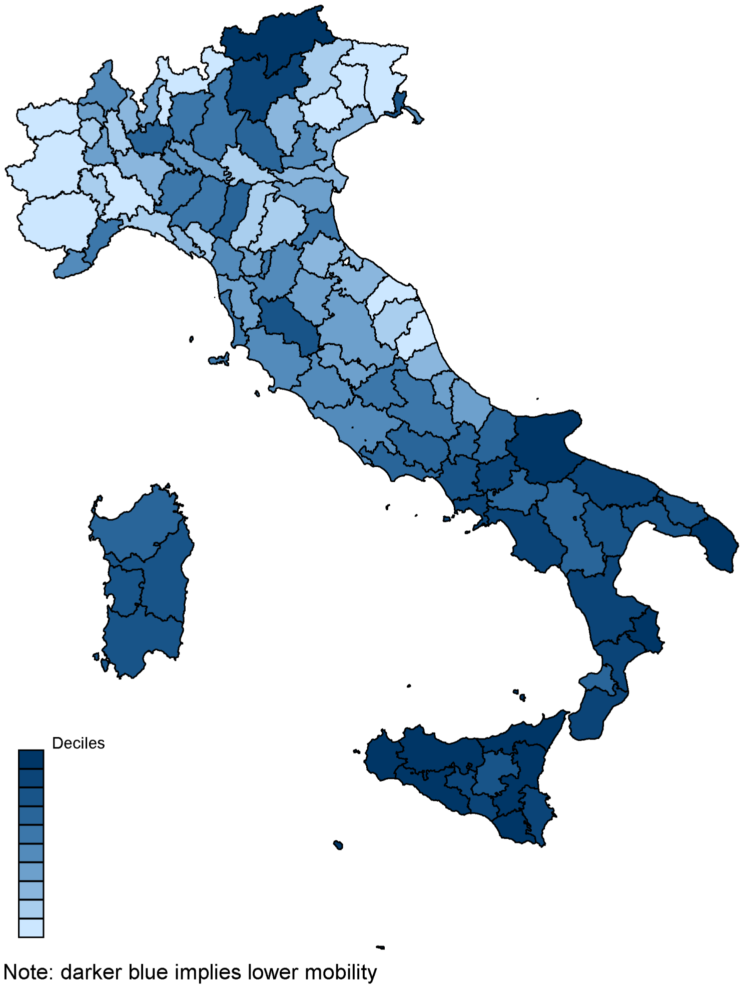


FIGURE 2. Social mobility (ICS-30) across Italian provinces

Table 6. ICS measures based on taxable income and local surnames: descriptive statistics

	N	Mean	St.Dev.	Percentiles 10	50	90
ICS based on taxable income, local	103	0.0243	0.0102	0.0124	0.0219	0.0399
ICS based on taxable income, local and tail 30	103	0.0507	0.0195	0.0326	0.0463	0.0721
ICS based on taxable income, local and tail 25	103	0.0546	0.0209	0.0340	0.0501	0.0747
ICS based on taxable income, local and tail 20	103	0.0587	0.0221	0.0385	0.0525	0.0826
ICS based on taxable income, local and tail 15	103	0.0643	0.0250	0.0414	0.0586	0.0895

Source: 2005 Italian tax records. Population: males aged 16-100.

Table 7. Pairwise correlations across ICS measures

	Full ICS	ICS-30	ICS-25	ICS-20	ICS-15	Local ICS	Local ICS-30	Local ICS-25	Local ICS-20	Local ICS-15
Full ICS	1.0000									
ICS-30	0.7010	1.0000								
ICS-25	0.6961	0.9960	1.0000							
ICS-20	0.6948	0.9941	0.9956	1.0000						
ICS-15	0.6908	0.9870	0.9893	0.9934	1.0000					
Local ICS	0.9077	0.5369	0.5299	0.5316	0.5339	1.0000				
Local ICS-30	0.6441	0.8805	0.8750	0.8739	0.8672	0.5779	1.0000			
Local ICS-25	0.6328	0.8737	0.8718	0.8698	0.8673	0.5679	0.9935	1.0000		
Local ICS-20	0.6150	0.8713	0.8693	0.8715	0.8721	0.5495	0.9875	0.9923	1.0000	
Local ICS-15	0.5745	0.8442	0.8436	0.8475	0.8548	0.5076	0.9686	0.9774	0.9849	1.0000

Notes: Full ICS refers to the ICS calculated with the full male population. All other ICS are calculated with the relevant tail of the surname distribution. Local ICS is calculated with only the 50% of the population with the most local surnames. Source: 2005 Italian tax records. Population: males aged 16-100.

to the most infrequent surnames. Overall, we again see marked variation across provinces and a monotonically increasing pattern of the ICS as we restrict to more and more infrequent surnames. The geographical breakdown of the local ICS provides a picture that is similar to the one that emerges from Figure 2.

Table 7 displays the pairwise correlations between all the ICS measures shown in Tables 5 and 6. Correlations are all very high and all significantly different from zero. We find particularly reassuring that the ICS measures based on local surnames correlate very strongly with their analogues based on both local and non-local surnames. This result suggests that differential migration patterns across provinces are unlikely to be a major concern.

ICS on aggregate province-level outcomes and controlling for provincial migration patterns.

5.1.1 Correlation between ICS and traditional measure of IM

In this section, we compare our ICS measure with a traditional measure of intergenerational mobility. For this comparison, we use the *Survey on Household Income and Wealth* (SHIW) from the Bank of Italy, which consists of repeated cross sections and includes some retrospective information on fathers' characteristics. This data set has been used by a number of studies to obtain measures of intergenerational mobility constructed on the basis of the traditional regression of children's outcomes on fathers' outcomes (Piraino, 2007; Mocetti, 2007; Checchi, Fiorio, and Leonardi, 2013).[22]

Unfortunately, given the limited sample size, the SHIW is not representative at the province level and codes for the province of residence are not distributed with the data. Hence, we can only calculate the traditional measure of mobility – following Checchi, Fiorio, and Leonardi (2013) – at the more aggregate level of the 20 Italian regions, which can be further aggregated into five macro-areas (North-West, North-East, Centre, South, Islands). We then also recalculate the ICS at the same geographical level (region or macro-area) and compare the two sets of indicators. Moreover, the retrospective information on fathers that is collected in the SHIW does not include income; hence, we can only compute the traditional intergenerational correlation coefficient from a regression of children's years of schooling on fathers' years of schooling.

The results are displayed in Table 8. Despite the small sample size and the rather different outcome indicators, our surname-based measure of IM and the traditional fathers-children coefficients are always positively correlated. The correlation is quite high when we focus on the five macro-areas (Table 8, top panel), although of course the very limited sample size does not allow the estimates to reach conventional levels of statistical significance. When we disaggregate results at the level of the 20 regions, correlations are still positive, though lower (Table 8, middle panel). This is not surprising because estimates of the traditional measures are based on smaller samples and thus are more imprecise. Yet, when we drop 25% of the regions with the least number of observations in the SHIW (these are very small regions with a small number of observations), the correlations become significant (Table 8, bottom panel).

[22] The only other data source used to estimate mobility in Italy is a survey conducted in 1985 on occupations with retrospective information on parents (Checchi, Ichino, and Rustichini, 1999).

Table 8. Pairwise correlations between ICS and father-son intergenerational correlation coefficient

	Full ICS	ICS-30	ICS-25	ICS-20	ICS-15
Father-son intergenerational correlation coefficient	0.7995	0.7301	0.7103	0.7007	0.7216
	(0.1045)	(0.1614)	(0.1788)	(0.1875)	(0.1688)
Level aggregation & observations	5 areas	5 areas	5 areas	5 areas	5 areas
Father-son intergenerational correlation coefficient	0.2544	0.2351	0.2398	0.2531	0.2685
	(0.2790)	(0.3185)	(0.3085)	(0.2816)	(0.2523)
Level aggregation & observations	20 regions	20 regions	20 regions	20 regions	20 regions
Father-son intergenerational correlation coefficient	0.4620	0.6647*	0.6836*	0.6792*	0.7070*
	(0.0830)	(0.0069)	(0.0050)	(0.0054)	(0.0032)
Level aggregation	20 regions	20 regions	20 regions	20 regions	20 regions
Observations (exclude 5 regions with least observations)	15 regions	15 regions	15 regions	15 regions	15 regions

Pairwise correlations and p-values in parentheses. (*) indicates significance at the 5% level or better. The father-son intergenerational correlation coefficients are computed as in Checchi, Fiorio, and Leonardi (2013). ICS measures as in Tables 5 and 6. Full ICS refers to the ICSs calculated with the full male population ICS. All other ICS are calculated with the relevant tail of the surname distribution.

Overall, these results are reassuring because they indicate that the ICSs are indeed capturing mobility patterns across geographical areas. We can, thus, confidently use our province-level ICS to explore how social mobility correlates with a number of meaningful macroeconomic outcomes.[23]

6 Intergenerational mobility and macroeconomic outcomes

We now turn to the analysis of the correlations between the ICS measures and our battery of macroeconomic outcomes. As we discussed in section 4.2, we organise these many outcomes in three groups. The first group (section 6.1) includes value added per capita, educational attainment, inequality and social capital. The second category (section 6.2) refers to labour market outcomes and the degree of trade openness in each province. The third group (section 6.3) includes, instead, a number socio-political outcomes such as life expectancy, suicide rates, crime rates and public sector activity.

[23] In a very recent paper Acciari, Polo, and Violante (2016) calculate income mobility measures for Italy using administrative data and standard parent-children regressions. The authors have correlated our ICS measures with theirs and find that the correlation is high and highly significant. In fact, their results are very consistent with ours. They find that intergenerational mobility is sharply heterogeneous across provinces, with a steeply positive South-North gradient. They also find that many good outcomes correlate with IGM.

6.1 Correlating Social Mobility and Key Variables

Table 9 presents the coefficients obtained from regressing the ICS-30 on the first group of outcomes.[24] Column 1 displays the coefficients from simple univariate regressions, in column 2 we add controls for value added per capita (when looking at other outcomes). In column 3 we add a North/South dummy and in column 4 we add controls for net migration flows.[25]

Recalling that a higher ICS implies lower mobility, the table shows that outcomes such as value added and social capital are consistently positively and significantly related to higher mobility; inequality, however, as measured from our tax data by the standard deviation of log(1+taxable income), and low education levels are related to lower mobility. This pattern also emerges consistently when controlling for value added per capita (column 2), a North/South dummy (column 3) or when controlling for migration flows (column 4), suggesting that the results are not driven by the well-known Italian North-South divide. Figures 3 and 4 show our regressions results graphically.[26] It is particularly noteworthy that the correlation between the 1981 value added and mobility is significant and negatively related to the ICS as the average between 1999 and 2004. Given that IM is presumably a very slow-moving process, this evidence hints at the fact that our correlations are not driven by some omitted variable simultaneously affecting both mobility and economic performance, as one would clearly worry when looking at our results using value added from the early 2000s. Our ICSs are measured almost 25 years

[24]Recall that the ICS-30 is calculated on male individuals whose surname contains at most 30 people. All of our results are robust to different definitions of the ICS. Results using the ICS calculated on all male individuals and results restricted to males with a local surname are in Appendix E. Results also including females and results excluding self-employed workers are in Appendix F. We refer the reader to the working paper version for results restricted to individuals whose surname contains less than 15, 20 and 25 people.

[25]Northern provinces are Alessandria, Aosta, Arezzo, Asti, Belluno, Bergamo, Biella, Bologna, Bolzano, Brescia, Como, Cremona, Cuneo, Ferrara, Firenze, Forli, Genova, Gorizia, Grosseto, Imperia, La Spezia, Lecco, Livorno, Lodi, Lucca, Mantova, Massa Carrara, Milano, Modena, Novara, Padova, Parma, Pavia, Piacenza, Pisa, Pistoia, Pordenone, Prato, Ravenna, Reggio Emilia, Rimini, Rovigo, Savona, Siena, Sondrio, Torino, Trento, Treviso, Trieste, Udine, Varese, Venezia, Verbania, Vercelli, Verona, Vicenza. Southern provinces are Agrigento, Ancona, Ascoli Piceno, Avellino, Bari, Benevento, Brindisi, Cagliari, Caltanissetta, Campobasso, Caserta, Catania, Catanzaro, Chieti, Cosenza, Crotone, Enna, Foggia, Frosinone, Isernia, Laquila, Latina, Lecce, Macerata, Matera, Messina, Napoli, Nuoro, Oristano, Palermo, Perugia, Pesaro Urbino, Pescara, Potenza, Ragusa, Reggio Calabria, Rieti, Roma, Salerno, Sassari, Siracusa, Taranto, Teramo, Terni, Trapani, Vibo Valentia, Viterbo.

[26]The province with the highest ICS is the province of Bolzano (see also Table C1). This is a region with two ethnic groups (i.e., Italian origin and Austrian origin). It is likely that in this context, surnames in Bolzano capture both family as well as ethnic information. A cleaner estimate of mobility for Bolzano based on the ICS should distinguish between Italian and German sounding surnames, which unfortunately we cannot do since we only have information on the coded surnames. However, we are reassured that our correlation results are not driven by the imprecision of the ICS in this one province.

Table 9. Relationship between the ICS-30 and key outcomes

	(1)	(2)	(3)	(4)
Economic activity				
Value added per capita (avg 1999-2004)	-0.030		-0.019	-0.011
	(0.004)***		(0.007)***	(0.006)**
Value added per capita (1981)	-0.042		-0.046	-0.026
	(0.005)***		(0.009)***	(0.009)***
Inequality				
Standard deviation of log income	0.037	0.038	0.046	0.025
	(0.004)***	(0.007)***	(0.008)***	(0.006)***
Schooling (lack of)				
Individuals aged 25-64 with at most 8 years of schooling	0.068	0.040	0.052	0.032
	(0.013)***	(0.013)***	(0.012)***	(0.013)**
Early school dropout aged 18-24	0.024	0.017	0.020	0.015
	(0.005)***	(0.004)***	(0.004)***	(0.004)***
Social capital				
Voter turnout (Chamber of Deputies)	-0.171	-0.124	-0.128	-0.063
	(0.023)***	(0.032)***	(0.029)***	(0.042)
Voter turnout (Senate of the Republic)	-0.084	-0.025	-0.033	0.012
	(0.021)***	(0.023)	(0.022)	(0.023)
Voter turnout (European Parliament)	-0.108	-0.077	-0.081	-0.052
	(0.015)***	(0.016)***	(0.015)***	(0.018)***
Newspaper sales per capita	-0.008	-0.004	-0.004	-0.003
	(0.002)***	(0.002)*	(0.002)**	(0.002)
Controls:				
Value added per capita	NO	YES	NO	NO
North/South dummy	NO	NO	YES	NO
Net migration flows (avg. 1999-2002)	NO	NO	NO	YES

Notes: Each coefficient is obtained from a separate regression of the ICS-30 on each variable. ICS-30 refers to the ICS calculated including only surnames that contain at most 30 people (see Section 5). The number of observations equals the number of provinces (103) in all regressions, except those that refer to 1981, when the number of provinces was equal to 95. Standard errors in parentheses. (***) indicates significance at the 1% level, (**) indicates significance at the 5% level and (*) indicates significance at the 10% level.

after 1981, thus we corroborate the findings in Chetty, Hendren, Kline, and Saez (2014), who find similar patterns using indicators of mobility and economic outcomes that are measured about 10 years apart.[27]

The relationship between intergenerational mobility and inequality indeed has a special interest on its own. A clear positive correlation between the intergenerational elasticity of earnings and the degree of cross-sectional inequality – named the "Great Gatsby Curve" – exists across countries.[28] This correlation has become the focus of a large public debate, which often

[27]The ICS30 for cohort aged 30 to 45 years old is very highly correlated with the ICS30 for the group 16 to 100 years old used in the paper.

[28]The curve was introduced in a 2012 speech by Alan Krueger, former chairman of the Council of Economic Advisers (Krueger, 2012) using data from Miles Corak (Corak, 2013a,b). The name was coined by former CEA staff economist Judd Cramer in reference to the upwardly mobile character in F. Scott Fitzgerald's novel. For more details about the origin of this curve, see Miles Corak's blog post at this link.

interprets it as the result of institutional differences: inequality and the prevalence of inheritance being low in countries with larger government intervention, as in the Nordic countries, and high in *laissez-faire* societies such as the Anglo-Saxon countries. The plot of the Italian Great Gatsby curve in Figure 4 clearly shows that in provinces where income inequality is lower, inheritance is less prevalent.[29] This result is noteworthy because all Italian provinces share the same institutional framework: intergenerational mobility correlates with low inequality even holding constant the institutional setup.

6.2 Correlating social mobility and other economic outcomes

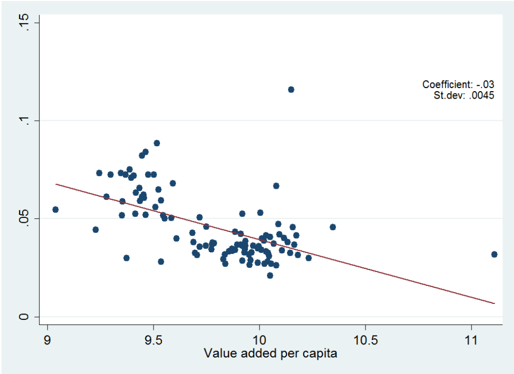
In Table 10 we report the correlations between the ICS and the second category of outcomes, namely labour market indicators and trade openness. Results clearly show that intergenerational mobility correlates positively with "good" economic outcomes, such as employment and openness, and negatively with "bad" economic outcomes, even after controlling for the level of economic activity (column 2), differences (observed and unobserved) between the North and the South of the country (column 3) or controlling for net migration flows (column 4).

6.3 Correlating social mobility and other socio-political outcomes

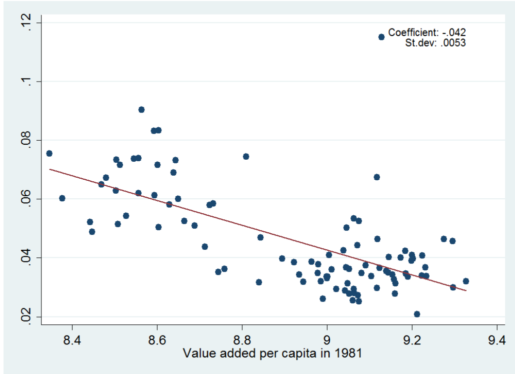
We now turn to our third category of outcomes, namely socio-political variables other than social capital. Results are presented in Table 11 using the same format of the previous tables. The outcomes are classified into four broad groups: indicators of life expectancy, crime rates, suicide rates and public sector activity.

Contrary to the results of the previous sections, the estimates in Table 11 do not seem to follow any clear path. Social mobility correlates with higher life expectancy for females, but not for males. There is some indication of a correlation with higher suicide rates, which nevertheless disappears when controlling for value added per capita, for the North/South dummy or migration flows. The same happens for crime: correlations are mostly, but not always, positive in column (1) and become largely insignificant in columns (2) and (3). Regarding the activity of the public sector, we find that higher mobility correlates negatively with the value of public

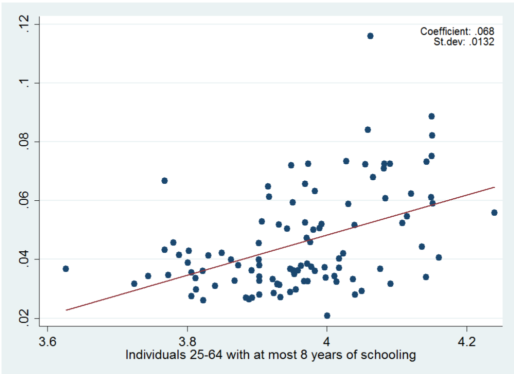
[29] The same qualitative result is found when using the Gini coefficient of income as an inequality measure.



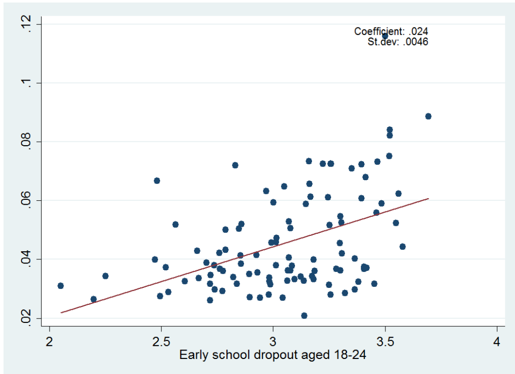
(a) value added per capita



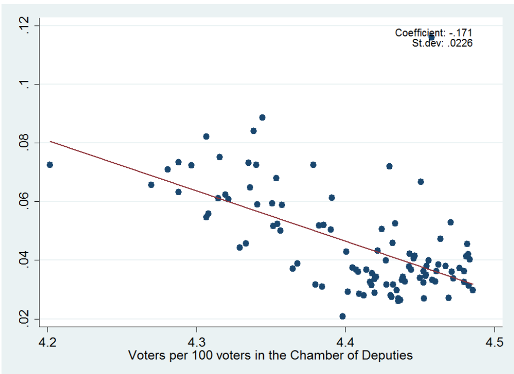
(b) value added per capita in 1981



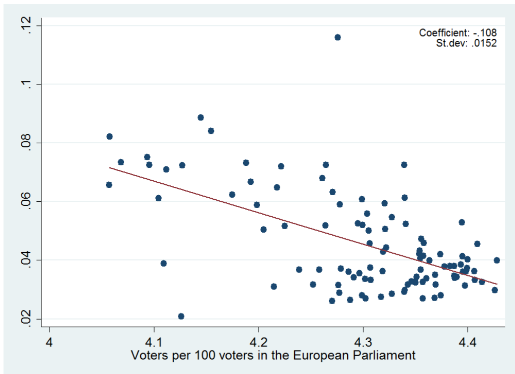
(c) Individuals 25-64 with at most 8 years of schooling



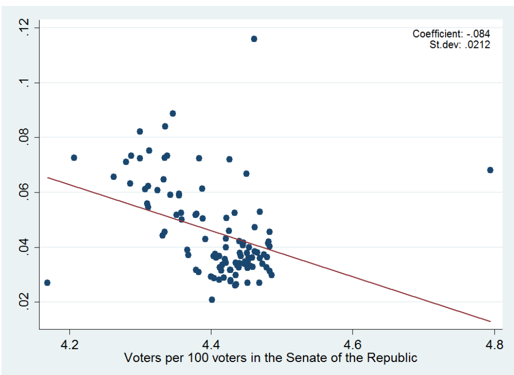
(d) Early school dropout aged 18-24



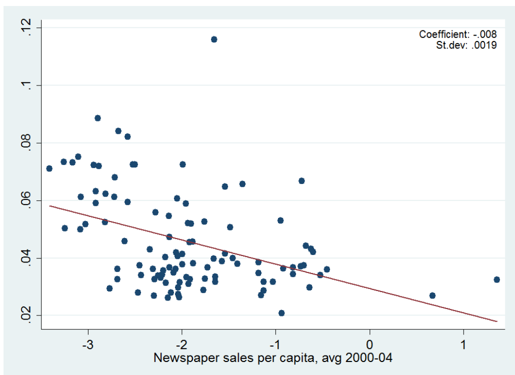
(e) Voter turnout in chamber of deputies elections



(f) Voter turnout in EU elections



(g) voter turnout in Senate of the Republic elections



(h) Newspaper sales per capita

FIGURE 3. Correlation between mobility (ICS-30) and key economic outcomes

Table 10. Relationship between the ICS-30 and other economic outcomes

	(1)	(2)	(3)	(4)
Economic activity				
Protested cheques per 1,000 inhabitants	0.009	0.004	0.001	0.003
	(0.002)***	(0.002)	(0.003)	(0.002)
Labour market outcomes				
Unemployment rate	0.015	0.013	0.014	0.008
	(0.002)***	(0.003)***	(0.003)***	(0.003)**
Unemployment rate (males)	0.013	0.012	0.013	0.007
	(0.002)***	(0.003)***	(0.003)***	(0.003)**
Unemployment rate (females)	0.016	0.014	0.014	0.008
	(0.002)***	(0.004)***	(0.003)***	(0.003)**
Unemployment rate (age 15-24)	0.014	0.008	0.009	0.004
	(0.002)***	(0.003)**	(0.003)***	(0.003)
Long-term unemployment rate (12 months or more)	0.010	0.006	0.007	0.003
	(0.002)***	(0.002)**	(0.002)***	(0.002)
Employment rate	-0.071	-0.066	-0.066	-0.041
	(0.009)***	(0.015)***	(0.014)***	(0.016)***
Employment rate (males)	-0.098	-0.059	-0.065	-0.024
	(0.014)***	(0.023)***	(0.022)***	(0.023)
Employment rate (females)	-0.045	-0.046	-0.045	-0.034
	(0.005)***	(0.008)***	(0.008)***	(0.009)***
Employment rate (age 15-24)	-0.024	-0.018	-0.021	-0.011
	(0.003)***	(0.005)***	(0.005)***	(0.005)**
Employment rate (high school, aged 25-64)	-0.102	-0.111	-0.113	-0.084
	(0.010)***	(0.018)***	(0.018)***	(0.020)***
Employment rate (college graduate, aged 25-64)	-0.155	-0.103	-0.122	-0.068
	(0.022)***	(0.032)***	(0.040)***	(0.029)**
Participation rate (age 15-64)	-0.106	-0.089	-0.095	-0.057
	(0.013)***	(0.021)***	(0.021)***	(0.022)**
Participation rate (age 15-64 males)	-0.159	-0.027	-0.028	0.040
	(0.037)***	(0.045)	(0.047)	(0.044)
Participation rate (age 15-64 females)	-0.055	-0.052	-0.053	-0.038
	(0.006)***	(0.009)***	(0.009)***	(0.010)***
Participation rate (age group 15-24)	-0.044	-0.024	-0.025	-0.013
	(0.007)***	(0.009)**	(0.010)**	(0.009)
Openness				
Imports to value added	-0.007	-0.002	-0.002	-0.001
	(0.002)***	(0.002)	(0.002)	(0.002)
Exports to value added	-0.009	-0.006	-0.007	-0.004
	(0.001)***	(0.002)***	(0.002)***	(0.002)***
Controls:				
Value added per capita	NO	YES	NO	NO
North/South dummy	NO	NO	YES	NO
Net migration flows (avg. 1999-2002)	NO	NO	NO	YES

Notes: Each coefficient is obtained from a separate regression of the ICS-30 on each variable. ICS-30 refers to the ICS calculated including only surnames that contain at most 30 people (see Section 5). Standard errors in parentheses. (***) indicates significance at the 1% level, (**) indicates significance at the 5% level and (*) indicates significance at the 10% level.

Table 11. Relationship between the ICS-30 and other socio-political outcomes

	(1)	(2)	(3)	(4)
Life expectancy				
Life expectancy at birth, males	-0.018 (0.154)	0.045 (0.130)	-0.110 (0.132)	0.182 (0.121)
Life expectancy at 65, males	0.048 (0.054)	0.026 (0.046)	-0.019 (0.048)	0.069 (0.042)
Life expectancy at birth, females	-0.635 (0.173)***	-0.301 (0.165)*	-0.416 (0.157)***	-0.080 (0.164)
Life expectancy at 65, females	-0.269 (0.051)***	-0.151 (0.054)***	-0.176 (0.051)***	-0.079 (0.055)
Crime Rates				
Total crimes	-0.013 (0.006)**	-0.003 (0.005)	-0.002 (0.006)	-0.007 (0.005)
Violent crimes	0.002 (0.006)	0.003 (0.005)	0.003 (0.005)	-0.002 (0.005)
Thefts	-0.010 (0.004)**	-0.001 (0.004)	0.000 (0.004)	-0.002 (0.003)
Other crimes	-0.006 (0.007)	-0.007 (0.006)	-0.003 (0.006)	-0.013 (0.005)**
Murders per 100,000 inhabitants	0.007 (0.002)***	0.002 (0.002)	0.004 (0.002)*	0.000 (0.002)
Petty thefts per 100,000 inhabitants	-0.006 (0.002)***	-0.001 (0.002)	-0.002 (0.002)	-0.002 (0.001)
Snatching per 100,000 inhabitants	0.004 (0.002)**	0.003 (0.001)**	0.003 (0.002)**	0.002 (0.001)
Burglaries per 100,000 inhabitants	-0.022 (0.004)***	-0.014 (0.004)***	-0.014 (0.004)***	-0.009 (0.004)**
Theft of parked cars per 100,000 inhabitants	-0.013 (0.003)***	-0.004 (0.003)	-0.005 (0.003)	-0.002 (0.003)
Car thefts per 100,000 inhabitants	0.006 (0.002)***	0.003 (0.002)	0.003 (0.002)	0.000 (0.002)
Scams per 100,000 inhabitants	-0.012 (0.004)***	-0.007 (0.004)*	-0.008 (0.004)**	-0.007 (0.004)**
Smuggling offences per 100,000 inhabitants	0.003 (0.001)***	0.002 (0.001)**	0.002 (0.001)**	0.001 (0.001)
Drug production and sale for 100,000 inhabitants	-0.004 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.002 (0.003)
Exploitation of prostitution per 100,000 inhabitants	-0.011 (0.002)***	-0.007 (0.002)***	-0.008 (0.002)***	-0.006 (0.002)***
Distraints per 1,000 inhabitants aged 18+	0.003 (0.003)	0.001 (0.003)	0.000 (0.003)	0.001 (0.002)
Distraints per 1,000 families	0.005 (0.003)	0.002 (0.003)	0.000 (0.003)	0.002 (0.002)
Suicides Rates				
Suicides per 100,000 - Total	-0.019 (0.004)***	-0.010 (0.004)***	-0.010 (0.004)**	-0.006 (0.004)
Suicides per 100,000 population - Males	-0.006 (0.002)***	-0.002 (0.002)	-0.001 (0.002)	0.000 (0.002)
Suicides per 100,000 population - Females	-0.007 (0.002)***	-0.003 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Suicide attempts per 100,000 - Total	-0.009 (0.003)***	-0.001 (0.003)	0.000 (0.003)	0.000 (0.003)
Suicide attempts per 100,000 - Males	-0.006 (0.002)***	-0.001 (0.002)	0.000 (0.002)	-0.001 (0.002)
Suicide attempts per 100,000 - Females	-0.003 (0.002)**	-0.001 (0.002)	0.000 (0.002)	-0.001 (0.001)
Public sector activity				
Value of public works started (pct VA)	0.004 (0.002)**	0.002 (0.002)	0.004 (0.002)**	0.004 (0.002)***
Value of public works started by provinces (pct VA)	0.005 (0.001)***	0.004 (0.001)***	0.004 (0.001)***	0.003 (0.001)***
Value of public works (construction sector, pct VA)	0.008 (0.002)***	0.006 (0.002)***	0.008 (0.002)***	0.006 (0.002)***
Value of public works completed (pct VA)	0.007 (0.003)**	0.000 (0.003)	0.004 (0.003)	0.002 (0.002)
Value of public works completed by provinces (pct VA)	0.006 (0.002)***	0.005 (0.002)***	0.005 (0.002)***	0.004 (0.001)***
Percentage politicians with at least secondary education	0.035 (0.014)**	0.019 (0.013)	0.025 (0.013)*	0.017 (0.012)
Ratio of paid to committed expenses	-0.035 (0.048)	-0.011 (0.041)	-0.005 (0.042)	-0.003 (0.038)
Deficit per capita in euro	0.001 (0.002)	0.002 (0.002)	0.002 (0.002)	0.001 (0.002)
Growth rate of deficit per capita in euro	0.001 (0.001)	0.000 (0.001)	0.001 (0.001)	0.000 (0.001)
Controls:				
Value added per capita	NO	YES	NO	NO
North/South dummy	NO	NO	YES	NO
Net migration flows (avg. 1999-2002)	NO	NO	NO	YES

Notes: Each coefficient is obtained from a separate regression of the ICS-30 on each variable. ICS-30 refers to the ICS calculated including only surnames that contain at most 30 people (see Section 5). Standard errors in parentheses. (***) indicates significance at the 1% level, (**) indicates significance at the 5% level and (*) indicates significance at the 10% level.

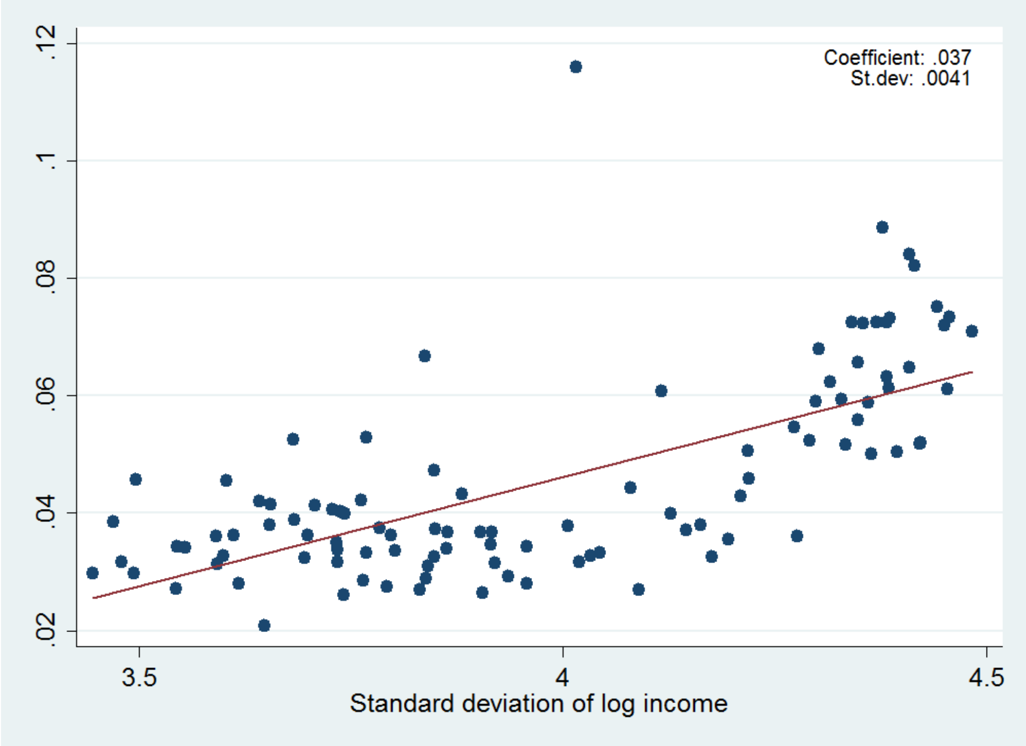


FIGURE 4. The Italian Great Gatsby Curve. Scatter plot of ICS-30 and inequality

works started and completed, and with a rough measure of the quality of local politicians (the proportion of politicians with at least a secondary education). Our data do not show any association with the ratio of paid to committed expenses nor with the local budget deficit (in both levels and growth rate).

Overall, no clear pattern emerges between intergenerational mobility and our array of social-political variables other than social capital. This result is perhaps not so surprising given that the interaction between mobility and these socio-political processes is presumably much more complex and unpredictable than the interaction with economic outcomes.

6.4 Summary

All our results for the different economic and socio-political outcomes can be visualised in Figure D1 in Appendix D which plots the value of the regression coefficients and their p-values for the key outcomes, the other economic outcomes and the other socio-political outcomes.

Appendices E and F provide similar figures for the different robustness checks that use altenative ICS measures and alternative population criteria.

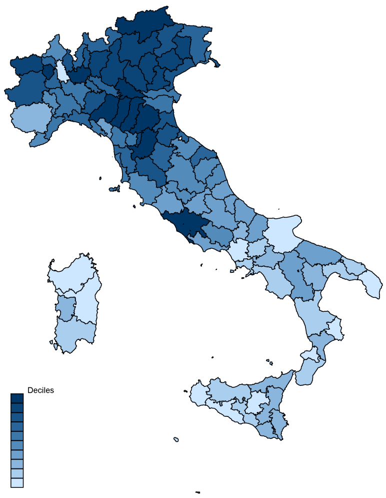
7 Conclusions

An important recent trend in the literature on intergenerational mobility investigates the correlation between indicators of social mobility and a variety of meaningful aggregate outcomes. Agreement in this area is still far from being reached. Chetty, Hendren, Kline, and Saez (2014) and Corak (2013b) find that social mobility differs across geographical areas and co-moves positively with economic activity and social capital, and co-moves negatively with inequality. Others, like Clark (2014), suggest that mobility is low and constant, and thus unrelated to aggregate variables.

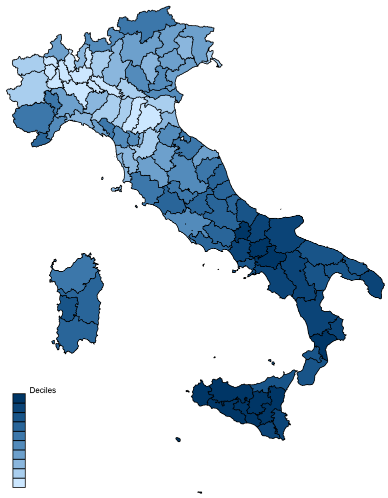
This paper uses within-country variation in social mobility and macroeconomic outcomes to contribute to this debate. We show that Italian provinces exhibit a large degree of variability in social mobility. This is particularly noteworthy in a centralised country like Italy, where the institutional framework is the same for all provinces. Thus, policies and political institutions are unlikely to be the main drivers of geographical differences in social mobility.

Our exercise shows that mobility correlates positively with economic activity, education and social capital, and negatively with inequality. Moreover, it correlates positively with all desirable economic outcomes and negatively with undesirable ones. The clear and systematic pattern that we document for economic outcomes and social capital does not emerge when we look at other socio-political variables.

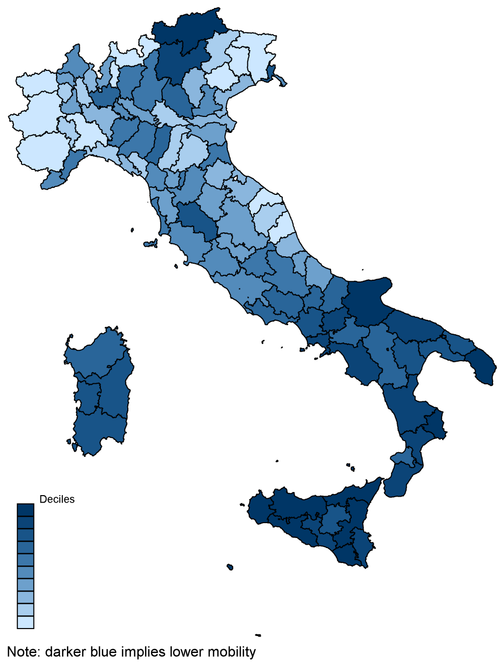
Although our approach does not allow us to make clear causal claims, we do improve over previous studies insofar as we can hold constant a vast number of institutional factors. We find that keeping constant institutions and policies, there are large differences in IM across provinces. Moreover, IM in Italian provinces correlates with aggregate outcomes in much the same manner as in Chetty, Hendren, Kline, and Saez (2014) for the United States. This necessarily implies that something beyond institutions and policies helps to shape IM and its relationship with aggregate outcomes. This, of course, does not mean that policies do not affect IM, but it does suggest a large degree of complexity in the socioeconomic equilibria that shape the workings of society.



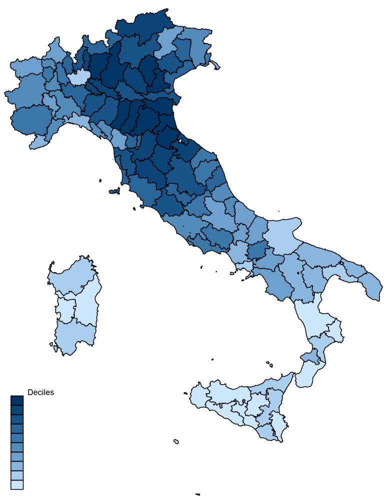
(a) value added per capita



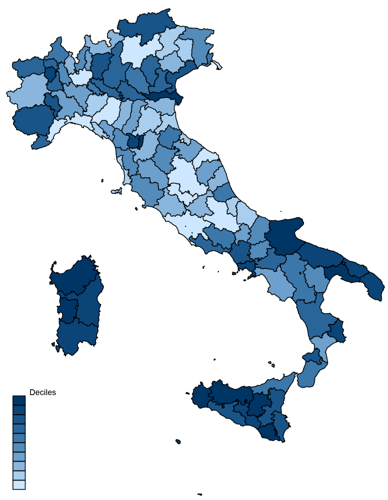
(b) inequality



(c) (inverse of) social mobility



(d) voters turnout



(e) % at most 8 years of schooling

FIGURE 5. Geographical patterns by province

References

Acciari, P., A. Polo, and G. L. Violante (2016). 'And yet, it moves': Intergenerational mobility in Italy. Mimeo New York University.

Anelli, M. and G. Peri (2013). Peer Gender Composition and Choice of College Major. NBER Working Papers 18744, National Bureau of Economic Research, Inc.

Aydemir, A. and H. Yazici (2015). Intergenerational education mobility and the level of development: Evidence from turkey. Mimeo, Sabanci University.

Barone, G. and S. Mocetti (2016). Intergenerational mobility in the very long run: Florence 1427-2011. Working Papers 1060, Bank of Italy.

Björklund, A., T. Eriksson, M. Jäntti, O. Raaum, and E. Österbacka (2002). Brother correlations in earnings in Denmark, Finland, Norway, and Sweden compared to the United States. *Journal of Population Economics 15*(4), 757–772.

Björklund, A. and M. Jäntti (1997). Intergenerational income mobility in Sweden compared to the United States. *American Economic Review 87*(5), 1009–1018.

Björklund, A. and K. G. Salvanes (2011). *Education and family background: Mechanisms and policies.* in E. Hanushek, S. Machin, and L. Woessmann (eds.), Handbook of the Economics of Education, Volume 3(3), pp. 201-247.

Black, S. E. and P. J. Devereux (2011). *Recent Developments in Intergenerational Mobility.* in Orley C. Ashenfelter and David Card (eds.), Handbook of Labor Economics, Volume 4B, Amsterdam: North-Holland, pp. 1487-1541.

Braga, M., M. Paccagnella, and M. Pellizzari (2016). The impact of college teaching on students' academic and labor market outcomes. *Journal of Labor Economics 34*(3), 781–822.

Checchi, D., C. V. Fiorio, and M. Leonardi (2013). Intergenerational persistence in educational attainment in Italy. *Economics Letters 118*, 229–232.

Checchi, D., A. Ichino, and A. Rustichini (1999). More equal but less mobile? Education financing and intergenerational mobility in Italy and in the U.S. *Journal of Public Economics 74*(3), 351–93.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? the geography of intergenerational mobility in the United States. *Quarterly Journal of Economics 129*(4), 1553–1623.

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility.* Princeton University Press. With Neil Cummins and Yu Hao and Daniel Diaz Vidal.

Collado, M. D., I. Ortuño-Ortín, and A. Romeo (2012). Long-run intergenerational social mobility and the distribution of surnames. Mimeo.

Comi, S. (2003). Intergenerational mobility in Europe: evidence from ECHP. Departmental Working Papers 2003-03, Department of Economics, Management and Quantitative Methods at Università degli Studi di Milano.

Corak, M. (2006). *Do Poor Children Become Poor Adults? Lessons for Public Policy from a Cross Country Comparison of Generational Earnings Mobility.* Research on Economic Inequality. Volume 13, Dynamics of Inequality, The Netherlands: Elsevier Press. Pages 143-188.

Corak, M. (2013a, Summer). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives 27*(3), 79–102.

Corak, M. (2013b). *Inequality from Generation to Generation: The United States in Comparison*. Robert Rycroft editor, The Economics of Inequality, Poverty, and Discrimination in the 21st Century, ABC-CLIO.

Couch, K. A. and T. A. Dunn (1997). Intergenerational correlations in labor market status: A comparison of the United States and Germany. *Journal of Human Resources 32*(1), 210–32.

Diaz Vidal, D. (2014). A surname analysis of social mobility and assortative mating in Chile, 1920-2004. Mimeo.

Fiorio, C. and F. D'Amuri (2005). Workers' tax evasion in italy. *Giornale degli Economisti e Annali di Economia 64*(2/3), 247–270.

Fox, W. R. and G. W. Lasker (1983). The distribution of surname frequencies. *International Statistical Review 51 (1)*, 81–87.

Gagliarducci, S. and T. Nannicini (2013, April). Do better paid politicians perform better? disentangling incentives from selection. *Journal of the European Economic Association 11*(2), 369–398.

Grawe, N. D. (2004). *Intergenerational mobility for whom? The experience of high- and low-earnings son in international perspective.* in Miles Corak (ed.), Generational Income Inequality, Cambridge University Press, pp.58-89.

Güell, M., M. Pellizzari, G. Pica, and J. V. Rodríguez Mora (2015). Correlating social mobility and economic outcomes. CEPR DP10496.

Güell, M., J. V. Rodríguez Mora, and C. I. Telmer (2007). Intergenerational mobility and the informative content of surnames. CEPR Discussion Paper No. 6316.

Güell, M., J. V. Rodríguez Mora, and C. I. Telmer (2014). The informational content of surnames, the evolution of intergenerational mobility and assortative mating. *The Review of Economic Studies*. DOI: 10.1093/restud/rdu041.

Heidrich, S. (2015). Intergenerational mobility in Sweden: a regional perspective. Umeå Economic Studies No.916, 2015.

Krueger, A. B. (2012). The rise and consequences of inequality in the United States. Speech of the Chairman of Council of Economic Advisers at the Center for American Progress on January 12th, 2012.

Mocetti, M. (2007). Intergenerational earnings mobility in Italy. *The B.E. Journal of Economic Analysis & Policy 7,*(Iss. 2 (Topics)), 1935–1682.

Mocetti, S. and E. Viviano (2015). Looking behind mortgage delinquencies. Working Papers 999, Bank of Italy.

Nybom, M. and K. Vosters (2017). Intergenerational persistence in latent socioeconomic status: Evidence from sweden. The Journal of Labor Economics, forthcoming.

Piraino, P. (2007). Comparable estimates of intergenerational income mobility in Italy. *The B.E. Journal of Economic Analysis & Policy 7,*(Iss. 2 (Topics)), 1935–1682.

Rubinstein, Y. and D. Brenner (2011). Pride and prejudice: Using ethnic-sounding names and inter-ethnic marriages to identify labor market discrimination. Mimeo, LSE.

Solon, G. (2016). What do we know so far about multigenerational mobility? *Economic Journal, forthcoming*.

Vosters, K. (2017). Is the simple law of mobility really a law? Testing Clark's hypothesis. The Economic Journal, forthcoming.

A Appendix: Mobility estimates using the Clark methodology

In recent work Gregory Clark assures to have found an "universal constant" of intergenerational mobility. Essentially, he correlates a surname-grouped average of income or wealth across generations for a certain cut of the population and reports an intergenerational elasticity of around 0.8, which he claims is essentially constant across societies and time. This intergenerational elasticity is much larger than anybody had previously measured and the lack of variation across societies and time is also at odds with previous literature. He then claims that this surname-grouped correlation reflects the persistence of an underlying "social status" variable that would be much more persistent than income or education, and that it is the "correct" variable to study.

In a nutshell, the surname-grouping methodology consists in averaging the income of all people sharing a surname in each of two cohorts of individuals, and then correlating the averages per surname of the younger and older cohort. It tries to mimic the standard parent-children regressions, but as all grouped estimators it suffers from a well known upward bias,[30] as *within-surname* mobility is not accounted for while persistent differences between surnames are made salient. Thus, for instance, if a surname average is systematically larger than the average of another surname (because, say, the first is prevalent in a rich province while the second is prevalent in a poor one), the correlation between older and younger cohorts is going to be very large. It might well be that the correlation between parents and children (not the surname average) is low because within each surname (and within each province) there is a large amount of mobility, but the surname-grouped correlation may still be very large.

In this appendix we replicate Clark's methodology on the Italian individual tax records used in this paper in order to investigate (*i*) if his results can be replicated and (*ii*) under which data restrictions this is possible. Given that Clark does not provide much information on how he selects his samples, we will experiment with different cuts of the Italian tax data. We confirm the finding of previous studies that have been unable to reproduce and/or explain Clark's results (see for instance Vosters (2017), Nybom and Vosters (2017), and section D of the Online Appendix of Chetty, Hendren, Kline, and Saez (2014)). Nevertheless, we shed some light into the possible causes behind his findings.

A.1 Clark's results can not be reproduced with Italian data

For every given surname s in any province, we calculate the average income separately for the individuals younger than 50 ($\bar{Y_s}^{young}$) and older than 50 ($\bar{Y_s}^{old}$). Like Clark, we then run a country-level OLS regression of $\log(\bar{Y_s}^{old})$ on $\log(\bar{Y_s}^{young})$. The resulting coefficients are the estimates of the pseudo-intergenerational elasticity (pseudo-IGE hereafter) obtained by grouping the population by surnames.

The first column of table A1 shows results from an unweighted regression and the second column from a regression in which each observation is weighted by the relative frequency of the surname. In both columns the estimated pseudo-IGE are one order of magnitude smaller than those reported by Clark. This result based on Italian tax data confirms what Chetty et al. (2014) find using US

[30]See for instance Solon (2016).

Table A1. Surname-grouped pseudo-IGE

	(1) Unweighted	(2) Weighted
constant	7.33***	6.34***
	(0.019)	(0.0017)
pseudo-IGE	0.2***	0.34***
	(0.0002)	(0.0002)
num. obs.	186318	186318
R^2	0.05	0.16

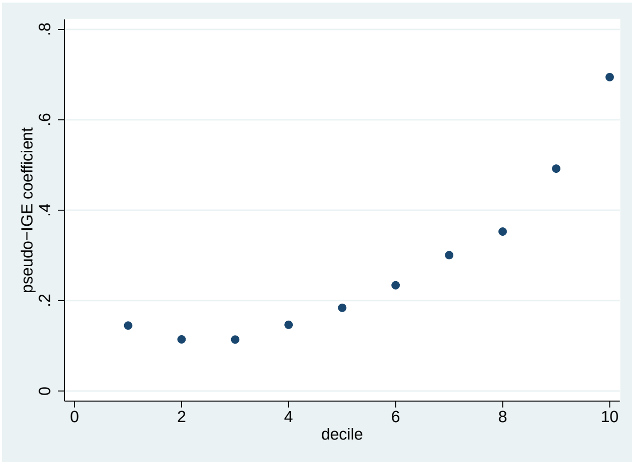


Figure A1. Surname-grouped pseudo-IGE per decile of surname frequency

administrative data. Using high-quality data of the whole population (and not arbitrarily chosen samples) the results that Clark reports can *not* be reproduced.

Notice that if we weight the surnames by their frequency, the estimated elasticity almost doubles (albeit it is still less than half of what Clark claims to find). This suggests that, like in Chetty et al. (2014), if we restrict the sample to frequent surnames, the correlation may increase.

This is indeed the case. In Figure A.1 we plot the pseudo-IGE obtained in a series of separate regressions in which we cut the population into deciles of surname frequency, i.e. on the extreme left of the horizontal axis we report the pseudo-IGE for the 10% most infrequent surnames, and on the extreme right for the 10% most frequent surnames. Figure A.1 shows that it is possible to obtain large values of the pseudo-IGE when restricting to very *frequent* surnames.

However, notice the irony: this is *exactly* the opposite of what Clark does. He focuses on very *infrequent* surnames, probably because that simplifies data collection and because within infrequent surnames family connections are likely to be stronger. Nevertheless, restricting to infrequent surnames, we obtain estimates of the pseudo-IGE that are much smaller than Clark's ones. This is true both using Italian tax data (as shown in the figure) and US administrative data: Chetty, Hendren, Kline, and Saez (2014) show that frequent surnames produce much larger surname-grouped IGE estimations than non-frequent surnames. They attribute this effect to the ethnic information that surnames contain, which is extremely reasonable in the US context as it is well known that surnames contain ethnic information.[31]

A.2 Surname-grouped correlations reflect mostly geographical differences in income

In the Italian context, the information that very common surnames most likely capture is the geographical origin. To understand whether this is indeed the case, we re-run our previous exercise at the *province* level. To do so, for every given surname s and province p, we calculate the average income separately for the individuals younger than 50 ($\bar{Y}_{s,p}^{young}$) and older than 50 ($\bar{Y}_{s,p}^{old}$). We then estimate the pseudo-IGE separately for each province.

To compare results, Figure A.2 adds to Figure A.1 the pseudo-IGE (averaged across provinces) for the same deciles of the distribution of surname frequencies used in the country-level exercise. Results show that, once we control for the geographic component, the differences between the pseudo-IGE estimated at different surname frequencies disappear, and a low pseudo-IGE is obtained for all frequencies: frequent surnames are like infrequent ones.

The reason is as follows. Surnames contain information about the geographic origin of their holder. Some surnames are very frequent in one province, and much less in other provinces. Italian provinces display persistent income differences, as the ranking of income per capita across Italian provinces moves slowly over time. The holders of a frequent surname are unlikely to be family related, and by the law of large numbers their average income is going to be very similar to the average income of the province. Thus, the average income of a frequent surname prevalent in a rich province is likely higher than the average income of a surname that is frequent in a poorer province, both for the old and for the young. The cross-province variation makes the correlation across generations very high. The within province variation is instead likely much smaller (in the limit zero) for frequent surnames as the average income of frequent surnames is very similar (in the limit identical) to the province average both for the old and the young, and is therefore captured by the constant in a province-level regression.[32]

Things are different for infrequent surnames because the average income of any infrequent surname does not reflect the average, persistent, income of the province. Thus, even when exploiting the cross-province variation correlations are not necessarily high, as they do not reflect province averages. The variation which is left within infrequent surnames is the the one induced by social mobility, chance, and noise added by mixing people with different degrees of family relatedness. All those things move the average income of the young relative to the average income of the old.

This is why, when we calculate the pseudo-IGE within a province, thus controlling for the geographical information embedded in frequent surnames, we are left – in all provinces and for all surname frequencies – with pseudo-IGE that are much lower than the ones that Clark purportedly has found

[31] Many papers have used the fact that surnames have ethnic information. For instance, Güell, Rodríguez Mora and Telmer (2007; 2014) show that surnames have ethnic information, and use this fact to control for ethnicity in Spain. Rubinstein and Brenner (2011) do the same for Israel.

[32] The pseudo-IGE is still identified in province-level regressions of frequent surnames because surnames that are frequent at the country-level are not equally frequent in all provinces. This generates the variation that identifies the parameter.

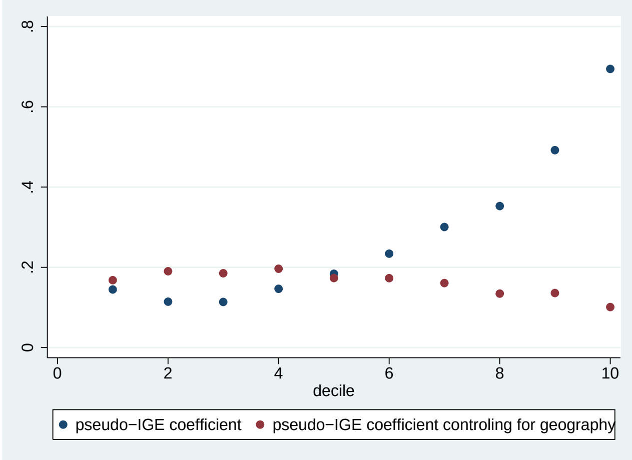


FIGURE A2. Surname-grouped pseudo-IGE per decile of surname frequency

across all societies and time periods.

Notice, though, that the pseudo-IGE, very much as the ICS, is *not* a direct measure of the intergenerational elasticity. They are both reflections of the "true" IGE, which can only be recovered either using a structural model (as in Güell, Rodríguez Mora, and Telmer (2014)) or accessing high-quality linked panels (as in Chetty et al. (2014) or Acciari et al. (2016)). In order to map pseudo-IGEs into IGEs we would need to model and understand how people is aggregated into surnames, their degree of family connections, etc.

The open question is how it is possible that Clark finds high pseudo-IGEs using infrequent surnames. One possibility is that pseudo-IGEs for infrequent surnames are very low when using high-quality administrative data covering the whole population; Clark, instead, uses selected samples. We have been unable to find cuts of the data that systematically produce high pseudo-IGE for infrequent surnames. Much more information on how Clark's sample selection procedure works, and how missing data are treated, is necessary in order to reproduce his results.

B Appendix: Macrodata sources

Table B1. Macroeconomic variables and years available

Variables	Years
Key outcomes	
Value added per capita	1981
Value added per capita	1999-2004
Individuals aged 25-64 with at most 8 years of schooling per 100 same age individuals	2004
Early school dropout aged 18-24 per 100 same age individuals	2004
Standard deviation of log income	2004
Voters turnout in the Chamber of Deputies per 100 voters	2006 and 2008
Voters turnout in the Senate of the Republic per 100 voters	2006 and 2008
Voters turnout in the European Parliament per 100 voters	2004
Newspaper sales per capita	2000-2004
Other economic outcomes	
Protested cheques[1] per 1,000 inhabitants	1999-2005
Unemployment rate	1999-2004
Unemployment rate - Males	1999-2004
Unemployment rate - Females	1999-2004
Unemployment rate (age 15-24)	1999-2004
Long-term unemployment rate (12 months or more)	2004
Employment rate	1999-2004
Employment rate - Males	1999-2004
Employment rate - Females	1999-2004
Employment rate (age 15-24)	1999-2004
Employment rate (high school, age 25-64)	2004
Employment rate (at least college graduate, age 25-64)	2004
Participation rate (age 15-64)	1999-2004
Participation rate (age 15-64) - Males	1999-2004
Participation rate (age 15-64) - Females	1999-2004
Participation rate (age 15-24)	1999-2004
Imports to value added	1999-2004
Exports to value added	1999-2004
Other socio-political outcomes	
Life expectancy at birth - Males	2002-2004
Life expectancy at 65 - Males	2002-2004
Life expectancy at birth - Females	2002-2004
Life expectancy at 65 - Females	2002-2004
Suicides per 100,000 inhabitants - Total	1999-2004
Suicides per 100,000 inhabitants - Males	1999
Suicides per 100,000 inhabitants - Females	1999
Suicide attempts per 100,000 inhabitants - Total	1999-2004
Suicide attempts per 100,000 inhabitants - Males	1999
Suicide attempts per 100,000 inhabitants - Females	1999
Total crimes	1999-2004
Violent crimes	2004
Thefts	1999-2004
Other crimes	1999-2003
Murders per 100,000 inhabitants	2004
Petty thefts per 100,000 inhabitants	1999-2004
Snatching per 100,000 inhabitants	1999-2004
Burglaries per 100,000 inhabitants	1999-2004
Theft of parked cars per 100,000 inhabitants	1999-2004
Car thefts per 100,000 inhabitants	1999-2004
Scams per 100,000 inhabitants	1999-2003
Smuggling offences per 100,000 inhabitants	1999-2003
Drug production and sale per 100,000 inhabitants	1999-2004
Exploitation of prostitution per 100,000 inhabitants	1999-2004
Distraints per 1,000 inhabitants aged 18 years and older	1999-2003
Distraints per 1,000 families	1999-2000 and 2003
Value of public works started (pct of VA)	2000
Value of public works started by Provincial institutions (pct of VA)	2000
Value of public works started in construction sector (pct of VA)	2000
Value of public works completed (pct of VA)	2000
Value of public works completed by Provincial institutions(pct of VA)	2000
Percentage politicians with at least secondary education	2001
Ratio of paid to committed expenses	2000-2004
Deficit per capita in Euros	1993-2004
Growth rate of deficit per capita in Euros ($\times$100)	1993-2004

Sources: Italian National Institute of Statistics (ISTAT) except *Value added per capita* in 1981 from *Istituto Gugliemo Tagliacarne*; *Standard deviation of log income* from 2005 Italian tax records; *Newspaper sales* from `dati.adsnotizie.it`; *Percentage of politicians with at least secondary education* from the Ministry of Internal Affairs. *Ratio of paid to committed expenses* and *Deficit per capita* from Gagliarducci and Nannicini (2013);

[1] Protested cheque (check/bill): formal notarial statement drawn up on behalf of a creditor and declaring that the debtor has dishonoured a bill of exchange or promissory note.

C Appendix: ICS mobility measures by provinces

This appendix provides the ICS measures used in the paper, see Table C1 below.

Table C1. Provinces ranked by ICS 30

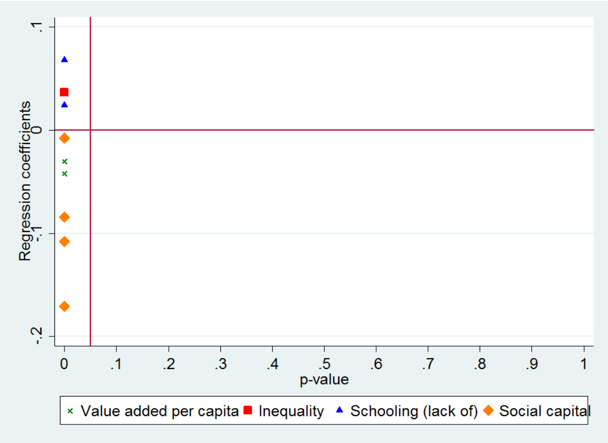
Province	Region	Area	ICS 30	Full ICS	ICS 30 local	ICS 30 (males and females)	ICS 30 (no self employed)
Aosta	Valle D'Aosta	North	0.021	0.014	0.022	0.024	0.016
Pordenone	Friuli Venezia Giulia	North	0.026	0.015	0.030	0.025	0.019
Sondrio	Lombardia	North	0.026	0.020	0.034	0.029	0.016
Ascoli Piceno	Marche	Centre	0.027	0.012	0.027	0.030	0.020
Treviso	Veneto	North	0.027	0.014	0.036	0.025	0.025
Lecco	Lombardia	North	0.027	0.022	0.052	0.020	0.034
Ancona	Marche	Centre	0.027	0.013	0.028	0.027	0.025
Cuneo	Piemonte	North	0.028	0.012	0.032	0.029	0.009
Torino	Piemonte	North	0.028	0.015	0.035	0.028	0.028
Alessandria	Piemonte	North	0.029	0.017	0.036	0.024	0.020
Udine	Friuli Venezia Giulia	North	0.029	0.016	0.035	0.025	0.022
Asti	Piemonte	North	0.029	0.016	0.031	0.025	0.006
Novara	Piemonte	North	0.030	0.024	0.037	0.030	0.022
Bologna	Emilia Romagna	North	0.030	0.017	0.045	0.031	0.036
Belluno	Veneto	North	0.031	0.023	0.035	0.033	0.017
Modena	Emilia Romagna	North	0.031	0.016	0.046	0.028	0.028
Massa Carrara	Toscana	Centre	0.032	0.019	0.029	0.031	0.037
Genova	Liguria	North	0.032	0.021	0.045	0.024	0.026
Biella	Piemonte	North	0.032	0.026	0.035	0.031	0.025
Macerata	Marche	Centre	0.032	0.016	0.033	0.031	0.019
Mantova	Lombardia	North	0.032	0.017	0.033	0.026	0.023
Forl'i	Emilia Romagna	North	0.033	0.013	0.031	0.034	0.014
Teramo	Abruzzo	South	0.033	0.015	0.029	0.041	0.029
Pavia	Lombardia	North	0.033	0.020	0.039	0.026	0.021
Rimini	Emilia Romagna	North	0.033	0.016	0.042	0.028	0.016
Pesaro Urbino	Marche	Centre	0.033	0.016	0.036	0.031	0.022
Venezia	Veneto	North	0.033	0.018	0.051	0.031	0.033
La Spezia	Liguria	North	0.034	0.027	0.037	0.025	0.027
Vicenza	Veneto	North	0.034	0.013	0.037	0.027	0.033
Rovigo	Veneto	North	0.034	0.020	0.040	0.024	0.024
Varese	Lombardia	North	0.034	0.026	0.045	0.028	0.033
Terni	Umbria	Centre	0.034	0.024	0.030	0.036	0.024
Vercelli	Piemonte	North	0.034	0.026	0.033	0.032	0.024
Perugia	Umbria	Centre	0.035	0.018	0.035	0.035	0.028
Pisa	Toscana	Centre	0.035	0.022	0.038	0.031	0.026
Pescara	Abruzzo	South	0.036	0.021	0.037	0.047	0.028
Cremona	Lombardia	North	0.036	0.023	0.041	0.027	0.031
Chieti	Abruzzo	South	0.036	0.018	0.033	0.042	0.027
Arezzo	Toscana	Centre	0.036	0.022	0.040	0.033	0.033
Como	Lombardia	North	0.036	0.028	0.058	0.027	0.033
Ferrara	Emilia Romagna	North	0.036	0.020	0.034	0.029	0.027
Pistoia	Toscana	Centre	0.037	0.021	0.043	0.029	0.030
Roma	Lazio	Centre	0.037	0.017	0.046	0.041	0.034
Lucca	Toscana	Centre	0.037	0.017	0.041	0.034	0.023
Imperia	Liguria	North	0.037	0.031	0.044	0.040	0.024
Padova	Veneto	North	0.037	0.016	0.047	0.033	0.033
Verbania	Piemonte	North	0.037	0.034	0.045	0.034	0.017
Grosseto	Toscana	Centre	0.038	0.027	0.043	0.028	0.030
Viterbo	Lazio	Centre	0.038	0.023	0.033	0.032	0.025
Firenze	Toscana	Centre	0.038	0.021	0.048	0.033	0.030
Lodi	Lombardia	North	0.038	0.029	0.043	0.026	0.041
Trieste	Friuli Venezia Giulia	North	0.039	0.035	0.051	0.037	0.021
Piacenza	Emilia Romagna	North	0.040	0.022	0.048	0.032	0.032
Rieti	Lazio	Centre	0.040	0.029	0.045	0.034	0.027
Brescia	Lombardia	North	0.040	0.017	0.052	0.035	0.043
Prato	Toscana	Centre	0.041	0.028	0.043	0.029	0.028
Ravenna	Emilia Romagna	North	0.041	0.015	0.051	0.034	0.030
Parma	Emilia Romagna	North	0.042	0.024	0.054	0.034	0.026
Bergamo	Lombardia	North	0.042	0.020	0.069	0.033	0.052
Livorno	Toscana	Centre	0.042	0.028	0.048	0.033	0.040
L'Aquila	Abruzzo	South	0.043	0.025	0.043	0.047	0.036
Savona	Liguria	North	0.043	0.036	0.057	0.031	0.028
Sassari	Sardegna	South	0.044	0.015	0.044	0.040	0.028
Reggio Emilia	Emilia Romagna	North	0.046	0.026	0.055	0.031	0.016
Milano	Lombardia	North	0.046	0.026	0.068	0.039	0.047
Frosinone	Lazio	Centre	0.046	0.022	0.053	0.048	0.030
Verona	Veneto	North	0.047	0.022	0.060	0.040	0.038
Potenza	Basilicata	South	0.050	0.028	0.054	0.056	0.039
Campobasso	Molise	South	0.050	0.034	0.048	0.057	0.039
Latina	Lazio	Centre	0.051	0.037	0.058	0.046	0.036
Vibo Valentia	Calabria	South	0.052	0.033	0.047	0.067	0.038
Isernia	Molise	South	0.052	0.035	0.043	0.058	0.039
Avellino	Campania	South	0.052	0.026	0.052	0.066	0.050
Brindisi	Puglia	South	0.052	0.024	0.053	0.055	0.052
Gorizia	Friuli Venezia Giulia	North	0.053	0.042	0.072	0.038	0.028
Siena	Toscana	Centre	0.053	0.038	0.047	0.040	0.050
Nuoro	Sardegna	South	0.055	0.018	0.047	0.046	0.028
Oristano	Sardegna	South	0.056	0.016	0.046	0.045	0.021
Caserta	Campania	South	0.059	0.027	0.065	0.070	0.044
Taranto	Puglia	South	0.059	0.029	0.062	0.057	0.061
Matera	Basilicata	South	0.059	0.044	0.055	0.063	0.049
Cagliari	Sardegna	South	0.061	0.018	0.067	0.057	0.053
Enna	Sicilia	South	0.061	0.039	0.060	0.072	0.065
Salerno	Campania	South	0.061	0.024	0.067	0.077	0.054
Napoli	Campania	South	0.062	0.020	0.069	0.074	0.047
Cosenza	Calabria	South	0.063	0.027	0.067	0.077	0.065
Catanzaro	Calabria	South	0.065	0.031	0.060	0.083	0.043
Reggio Calabria	Calabria	South	0.066	0.039	0.058	0.073	0.050
Trento	Trentino Alto Adige	North	0.067	0.031	0.122	0.052	0.019
Bari	Puglia	South	0.068	0.024	0.063	0.077	0.067
Caltanissetta	Sicilia	South	0.071	0.043	0.067	0.075	0.075
Benevento	Campania	South	0.072	0.045	0.078	0.079	0.072
Siracusa	Sicilia	South	0.072	0.039	0.077	0.069	0.061
Lecce	Puglia	South	0.072	0.024	0.072	0.077	0.058
Crotone	Calabria	South	0.072	0.040	0.077	0.087	0.044
Messina	Sicilia	South	0.073	0.033	0.074	0.077	0.086
Foggia	Puglia	South	0.073	0.035	0.071	0.075	0.067
Agrigento	Sicilia	South	0.073	0.033	0.066	0.083	0.073
Trapani	Sicilia	South	0.075	0.025	0.068	0.077	0.071
Palermo	Sicilia	South	0.082	0.031	0.084	0.095	0.101
Catania	Sicilia	South	0.084	0.035	0.086	0.084	0.099
Ragusa	Sicilia	South	0.089	0.034	0.101	0.092	0.071
Bolzano[1]	Trentino Alto Adige	North	0.116	0.053	0.151	0.082	0.051

Source: 2005 Italian tax records. Note: (1) The province with the highest ICS is Bolzano. Surnames in Bolzano are likely to capture both family as well as ethnic information (Italian origin vs. Austrian origin). See footnote 26 for further discussion.

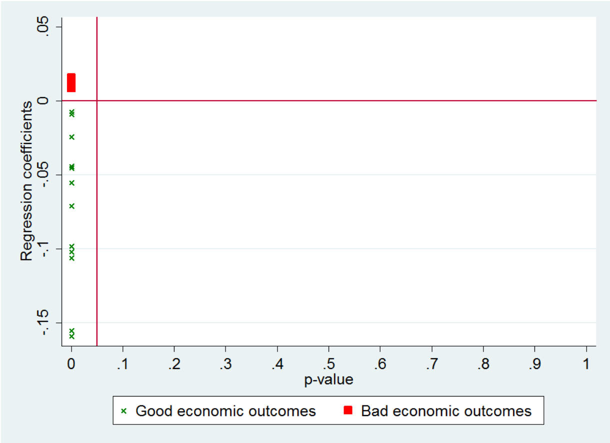
D Appendix: Summary graphs of the main results

Figure D1 summarises the correlations between the ICS and the macroeconomic variables by plotting the value of the regression coefficients and their p-values displayed in column 1 of Tables 9, 10 and 11, respectively.

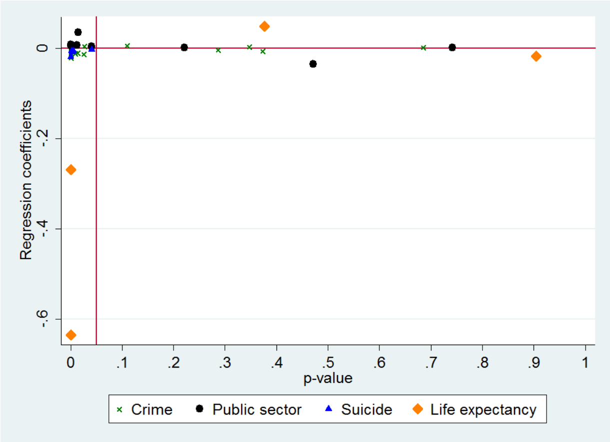
In order to provide a visual representation of this result, we classify all the variables into good and bad outcomes (see notes in Figure D1). It is clear from those graphs that mobility is high in places where economic outcomes are good.



(a) Key outcomes



(b) Other economic outcomes



(c) Other socio-political outcomes

FIGURE D1. Coefficients and p-values from separate regressions of the ICS-30 on province-level outcomes Notes: Good economic outcomes include: *Employment rate* and *Participation rate* for the different population groups, *Imports to value added* and *Exports to value added*. Bad economic outcomes include: *Protested cheques per 1000 inhabitants* and *Unemployment rate* for the different population groups (see Table B1).

E Appendix: Robustness checks using alternative ICS measures

This appendix provides results using alternative ICS measures. Tables E1-E3 and Figure E1 show results using the ICS calculated on the full population of males aged 16-100 (i.e., not restricted to individuals whose surname contains less than 30 people). Tables E4-E6 and Figure E2 show results further restricting our baseline ICS to the most local surnames in each province, as described in Section 2.2. The results presented in the body of the paper carry over to these alternative ICS measures.

TABLE E1. Relationship between the full ICS and key outcomes

	(1)	(2)	(3)	(4)
Economic activity				
Value added per capita (avg 1999-2004)	-0.009		-0.008	-0.011
	(0.003)***		(0.004)**	(0.006)**
Value added per capita (1981)	-0.012		-0.016	-0.026
	(0.003)***		(0.005)***	(0.009)***
Inequality				
Standard deviation of log income	0.011	0.012	0.019	0.025
	(0.002)***	(0.004)***	(0.005)***	(0.006)***
Schooling (lack of)				
Individuals aged 25-64 with at most 8 years of schooling	0.013	0.003	0.009	0.032
	(0.007)*	(0.008)	(0.007)	(0.013)**
Early school dropout (age 18-24)	0.005	0.003	0.004	0.015
	(0.003)*	(0.003)	(0.002)	(0.004)***
Social capital				
Voter turnout (Chamber of Deputies)	-0.059	-0.052	-0.060	-0.063
	(0.013)***	(0.019)***	(0.017)***	(0.042)
Voter turnout (Senate of the Republic)	-0.032	-0.017	-0.023	0.012
	(0.011)***	(0.013)	(0.013)*	(0.023)
Voter turnout (European Parliament)	-0.043	-0.037	-0.040	-0.052
	(0.008)***	(0.009)***	(0.009)***	(0.018)***
Newspaper sales per capita	-0.002	-0.001	-0.001	-0.003
	(0.001)**	(0.001)	(0.001)	(0.002)
Controls:				
Value added per capita	NO	YES	NO	NO
North/South dummy	NO	NO	YES	NO
Net migration flows (avg. 1999-2002)	NO	NO	NO	YES

Notes: each coefficient is obtained from a separate regression of the full ICS on each variable. Full ICS refers to the ICS calculated on the entire distribution of surnames. The number of observations equals the number of provinces (103) in all regressions, except those that refer to 1981, when the number of provinces was equal to 95. Standard errors in parentheses. (***) indicates significance at the 1% level, (**) indicates significance at the 5% level and (*) indicates significance at the 10% level.

Table E2. Relationship between the full ICS and other economic outcomes

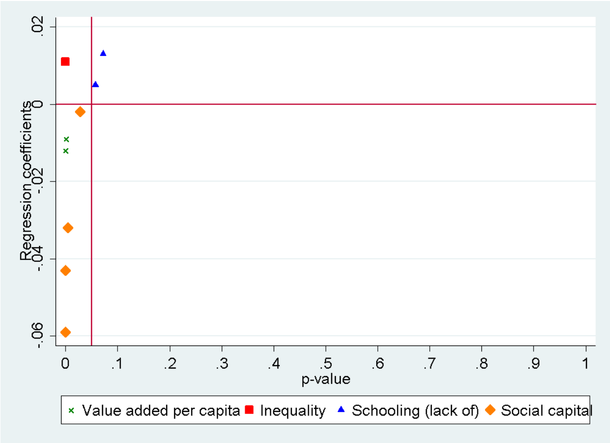
	(1)	(2)	(3)	(4)
Economic activity				
Protested cheques per 1,000 inhabitants	0.003	0.001	0.001	0.003
	(0.001)**	(0.001)	(0.002)	(0.002)
Labour market outcomes				
Unemployment rate	0.005	0.005	0.007	0.008
	(0.001)***	(0.002)***	(0.002)***	(0.003)**
Unemployment rate - Males	0.004	0.004	0.006	0.007
	(0.001)***	(0.002)***	(0.002)***	(0.003)**
Unemployment rate - Females	0.005	0.005	0.007	0.008
	(0.001)***	(0.002)**	(0.002)***	(0.003)**
Unemployment rate in the age group 15-24 years	0.005	0.005	0.007	0.004
	(0.001)***	(0.002)***	(0.002)***	(0.003)
Long-term unemployment rate (12 months or more) - Total	0.004	0.003	0.004	0.003
	(0.001)***	(0.001)**	(0.001)***	(0.002)
Employment rate	-0.027	-0.036	-0.041	-0.041
	(0.005)***	(0.008)***	(0.008)***	(0.016)***
Employment rate - Males	-0.042	-0.050	-0.058	-0.024
	(0.008)***	(0.012)***	(0.012)***	(0.023)
Employment rate - Females	-0.015	-0.020	-0.023	-0.034
	(0.003)***	(0.005)***	(0.005)***	(0.009)***
Employment rate aged 15-24	-0.009	-0.010	-0.014	-0.011
	(0.002)***	(0.003)***	(0.003)***	(0.005)**
Employment rate (high school aged 25-64)	-0.034	-0.042	-0.052	-0.084
	(0.007)***	(0.011)***	(0.011)***	(0.020)***
Employment rate of at least (college graduate aged 25-64)	-0.050	-0.038	-0.069	-0.068
	(0.013)***	(0.019)**	(0.023)***	(0.029)**
Participation rate (age 15-64)	-0.034	-0.034	-0.045	-0.057
	(0.008)***	(0.013)***	(0.012)***	(0.022)**
Participation rate (age 15-64) - Males	-0.066	-0.040	-0.057	0.040
	(0.019)***	(0.025)	(0.026)**	(0.044)
Participation rate (age 15-64) - Females	-0.017	-0.016	-0.021	-0.038
	(0.004)***	(0.006)***	(0.006)***	(0.010)***
Participation rate (age 15-24)	-0.021	-0.022	-0.029	-0.013
	(0.004)***	(0.005)***	(0.005)***	(0.009)
Trade Openness				
Imports to value added	-0.002	-0.001	-0.002	-0.001
	(0.001)**	(0.001)	(0.001)	(0.002)
Exports to value added	-0.003	-0.003	-0.003	-0.004
	(0.001)***	(0.001)***	(0.001)***	(0.002)***
Controls:				
Value added per capita	NO	YES	NO	NO
North/South dummy	NO	NO	YES	NO
Net migration flows (avg. 1999-2002)	NO	NO	NO	YES

Notes: Each coefficient is obtained from a separate regression of the full ICS on each variable. Full ICS refers to the ICS calculated on the entire distribution of surnames. Standard errors in parentheses. (***) indicates significance at the 1% level, (**) indicates significance at the 5% level and (*) indicates significance at the 10% level.

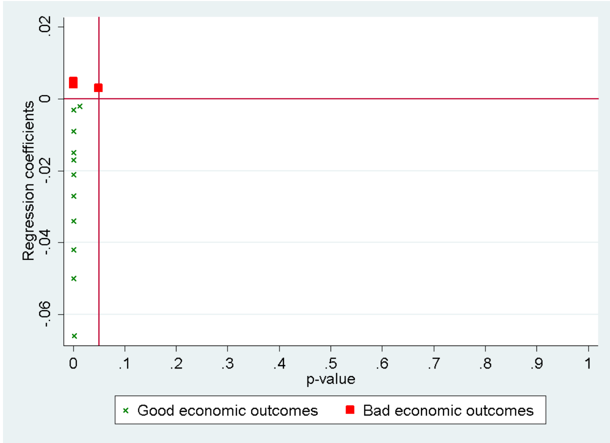
Table E3. Relationship between the full ICS and other socio-political outcomes

	(1)	(2)	(3)	(4)
Life expectancy				
Life expectancy at birth, males	-0.101	-0.083	-0.124	0.182
	(0.077)	(0.074)	(0.075)	(0.121)
Life expectancy at 65, males	-0.011	-0.018	-0.029	0.069
	(0.027)	(0.026)	(0.027)	(0.042)
Life expectancy at birth, females	-0.367	-0.297	-0.329	-0.080
	(0.085)***	(0.091)***	(0.087)***	(0.164)
Life expectancy at 65, females	-0.116	-0.093	-0.103	-0.079
	(0.027)***	(0.030)***	(0.029)***	(0.055)
Crime Rates				
Total crimes	-0.007	-0.005	-0.005	-0.007
	(0.003)**	(0.003)	(0.003)	(0.005)
Violent crimes	-0.004	-0.004	-0.004	-0.002
	(0.003)	(0.003)	(0.003)	(0.005)
Thefts	-0.007	-0.005	-0.006	-0.002
	(0.002)***	(0.002)**	(0.002)***	(0.003)
Other crimes	0.002	0.002	0.003	-0.013
	(0.003)	(0.003)	(0.003)	(0.005)**
Murders per 100,000 inhabitants	0.004	0.003	0.003	0.000
	(0.001)***	(0.001)**	(0.001)***	(0.002)
Petty thefts per 100,000 inhabitants	-0.002	-0.001	-0.001	-0.002
	(0.001)**	(0.001)	(0.001)	(0.001)
Snatching per 100,000 inhabitants	0.000	-0.001	-0.001	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Burglaries per 100,000 inhabitants	-0.009	-0.008	-0.010	-0.009
	(0.002)***	(0.002)***	(0.002)***	(0.004)**
Theft of parked cars per 100,000 inhabitants	-0.006	-0.005	-0.006	-0.002
	(0.001)***	(0.002)***	(0.002)***	(0.003)
Car thefts per 100,000 inhabitants	0.000	-0.001	-0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.002)
Scams per 100,000 inhabitants	-0.004	-0.003	-0.003	-0.007
	(0.002)*	(0.002)	(0.002)	(0.004)**
Smuggling offences per 100,000 inhabitants	0.001	0.001	0.001	0.001
	(0.001)***	(0.001)**	(0.001)**	(0.001)
Drug production and sale per 100,000 inhabitants	0.000	0.001	0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.003)
Exploitation of prostitution per 100,000 inhabitants	-0.004	-0.003	-0.004	-0.006
	(0.001)***	(0.001)***	(0.001)***	(0.002)***
Distraints per 1,000 inhabitants aged 18+	0.001	0.001	0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Distraints per 1,000 families	0.001	0.000	0.000	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Suicides Rates				
Suicides per 100,000 - Total	-0.005	-0.001	-0.002	-0.006
	(0.002)**	(0.002)	(0.002)	(0.004)
Suicides per 100,000 population - Males	0.000	0.002	0.002	0.000
	(0.001)	(0.001)*	(0.001)*	(0.002)
Suicides per 100,000 population - Females	-0.001	0.001	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)
Suicide attempts per 100,000 - Total	0.000	0.002	0.002	0.000
	(0.002)	(0.002)	(0.002)	(0.003)
Suicide attempts per 100,000 - Males	0.000	0.002	0.002	-0.001
	(0.001)	(0.001)*	(0.001)*	(0.002)
Suicide attempts per 100,000 - Females	0.000	0.001	0.001	-0.001
	(0.001)	(0.001)*	(0.001)	(0.001)
Public sector activity				
Value of public works started (pct VA)	0.002	0.002	0.002	0.004
	(0.001)**	(0.001)	(0.001)**	(0.002)***
Value of public works started by provinces (pct VA)	0.001	0.001	0.001	0.003
	(0.001)**	(0.001)*	(0.001)*	(0.001)***
Value of public works started (construction sector, pct VA)	0.002	0.001	0.002	0.006
	(0.001)*	(0.001)	(0.001)*	(0.002)***
Value of public works completed (pct VA)	0.004	0.003	0.004	0.002
	(0.001)***	(0.002)*	(0.002)***	(0.002)
Value of public works completed by provinces (pct VA)	0.002	0.002	0.002	0.004
	(0.001)**	(0.001)**	(0.001)**	(0.001)***
Percentage politicians with at least secondary education	0.002	-0.003	0.000	0.017
	(0.007)	(0.007)	(0.007)	(0.012)
Ratio of paid to committed expenses	0.012	0.019	0.019	-0.003
	(0.024)	(0.023)	(0.024)	(0.038)
Deficit per capita in euros	-0.001	0.000	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.002)
Growth rate of deficit per capita in euros	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Controls:				
Value added per capita	NO	YES	NO	NO
North/South dummy	NO	NO	YES	NO
Net migration flows (avg. 1999-2002)	NO	NO	NO	YES

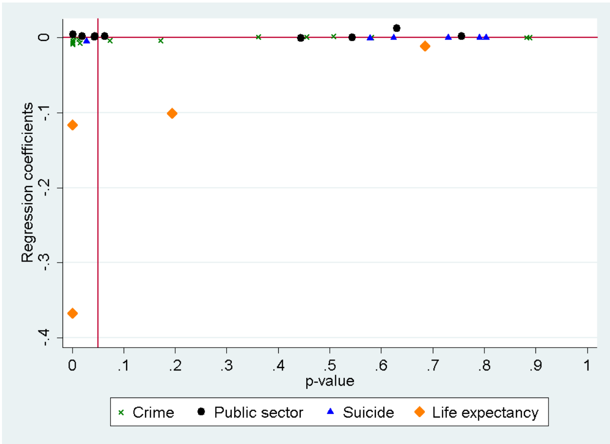
Notes: Each coefficient is obtained from a separate regression of the full ICS on each variable. Full ICS refers to the ICS calculated on the entire distribution of surnames. Standard errors in parentheses. (***) indicates significance at the 1% level, (**) indicates significance at the 5% level and (*) indicates significance at the 10% level.



(a) Key outcomes



(b) Other economic outcomes



(c) Other socio-political outcomes

FIGURE E1. Coefficients and p-values from separate regressions of the full ICS on province-level outcomes

Table E4. Relationship between the Local ICS-30 and key outcomes

	(1)	(2)	(3)	(4)
Economic activity				
Value added per capita (avg 1999-2004)	-0.016	-0.016	-0.012	-0.011
	(0.006)***	(0.006)***	(0.009)	(0.006)**
Value added (1981)	-0.025	-0.025	-0.038	-0.026
	(0.007)***	(0.007)***	(0.013)***	(0.009)***
Inequality				
Standard deviation of log income	0.024	0.032	0.043	0.025
	(0.006)***	(0.010)***	(0.011)***	(0.006)***
Schooling (lack of)				
Individuals aged 25-64 with at most 8 years of schooling	0.048	0.035	0.041	0.032
	(0.016)***	(0.018)*	(0.017)**	(0.013)**
Early school dropout (age 18-24)	0.022	0.019	0.020	0.015
	(0.005)***	(0.006)***	(0.005)***	(0.004)***
Social capital				
Voter turnout (Chamber of Deputies)	-0.114	-0.111	-0.112	-0.063
	(0.030)***	(0.043)**	(0.039)***	(0.042)
Voter turnout (Senate of the Republic)	-0.052	-0.023	-0.030	0.012
	(0.025)**	(0.030)	(0.029)	(0.023)
Voter turnout (European Parliament)	-0.095	-0.089	-0.090	-0.052
	(0.019)***	(0.022)***	(0.021)***	(0.018)***
Newspaper sales per capita	-0.006	-0.003	-0.004	-0.003
	(0.002)**	(0.003)	(0.003)	(0.002)
Controls:				
Value added per capita	NO	YES	NO	NO
North/South dummy	NO	NO	YES	NO
Net migration flows (avg. 1999-2002)	NO	NO	NO	YES

Notes: Each coefficient is obtained from a separate regression of the Local ICS-30 on each variable. Local ICS-30 refers to the ICS calculated including only surnames that contain at most 30 people (see Section 5) and are local (see Section 2.2). The number of observations equals the number of provinces (103) in all regressions, except those that refer to 1981, when the number of provinces was equal to 95. Standard errors in parentheses. (***) indicates significance at the 1% level, (**) indicates significance at the 5% level and (*) indicates significance at the 10% level.

Table E5. Relationship between the Local ICS-30 and other economic outcomes

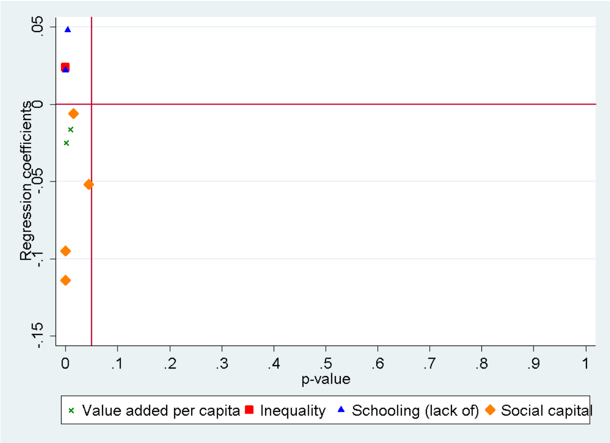
	(1)	(2)	(3)	(4)
Economic activity				
Protested cheques per 1,000 inhabitants	0.005	0.002	0.001	0.003
	(0.003)*	(0.003)	(0.004)	(0.002)
Labour market outcomes				
Unemployment rate	0.009	0.009	0.010	0.008
	(0.003)***	(0.005)*	(0.004)**	(0.003)**
Unemployment rate - Males	0.008	0.009	0.010	0.007
	(0.002)***	(0.004)**	(0.004)***	(0.003)**
Unemployment rate - Females	0.009	0.009	0.010	0.008
	(0.003)***	(0.005)*	(0.005)**	(0.003)**
Unemployment rate in the age group 15-24 years	0.007	0.003	0.004	0.004
	(0.003)**	(0.004)	(0.004)	(0.003)
Long-term unemployment rate (12 months or more) - Total	0.005	0.003	0.004	0.003
	(0.002)**	(0.003)	(0.003)	(0.002)
Employment rate	-0.042	-0.046	-0.050	-0.041
	(0.012)***	(0.021)**	(0.019)**	(0.016)***
Employment rate - Males	-0.048	-0.022	-0.032	-0.024
	(0.019)**	(0.031)	(0.029)	(0.023)
Employment rate - Females	-0.029	-0.039	-0.039	-0.034
	(0.007)***	(0.012)***	(0.011)***	(0.009)***
Employment rate aged 15-24	-0.014	-0.014	-0.018	-0.011
	(0.004)***	(0.007)**	(0.007)**	(0.005)**
Employment rate (high school aged 25-64)	-0.069	-0.103	-0.109	-0.084
	(0.015)***	(0.026)***	(0.025)***	(0.020)***
Employment rate of at least (college graduate aged 25-64)	-0.085	-0.060	-0.088	-0.068
	(0.029)***	(0.043)	(0.054)	(0.029)**
Participation rate (age 15-64)	-0.071	-0.086	-0.095	-0.057
	(0.018)***	(0.029)***	(0.028)***	(0.022)**
Participation rate (age 15-64) - Males	-0.067	0.017	0.008	0.040
	(0.045)	(0.059)	(0.061)	(0.044)
Participation rate (age 15-64) - Females	-0.040	-0.054	-0.056	-0.038
	(0.008)***	(0.013)***	(0.013)***	(0.010)***
Participation rate (age 15-24)	-0.025	-0.016	-0.019	-0.013
	(0.009)***	(0.013)	(0.014)	(0.009)
Trade Openness				
Imports to value added	-0.003	0.000	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Exports to value added	-0.006	-0.005	-0.006	-0.004
	(0.002)***	(0.002)**	(0.002)***	(0.002)***
Controls:				
Value added per capita	NO	YES	NO	NO
North/South dummy	NO	NO	YES	NO
Net migration flows (avg. 1999-2002)	NO	NO	NO	YES

Notes: Each coefficient is obtained from a separate regression of the Local ICS-30 on each variable. Local ICS-30 refers to the ICS calculated including only surnames that contain at most 30 people (see Section 5) and are local (see Section 2.2). Standard errors in parentheses. (***) indicates significance at the 1% level, (**) indicates significance at the 5% level and (*) indicates significance at the 10% level.

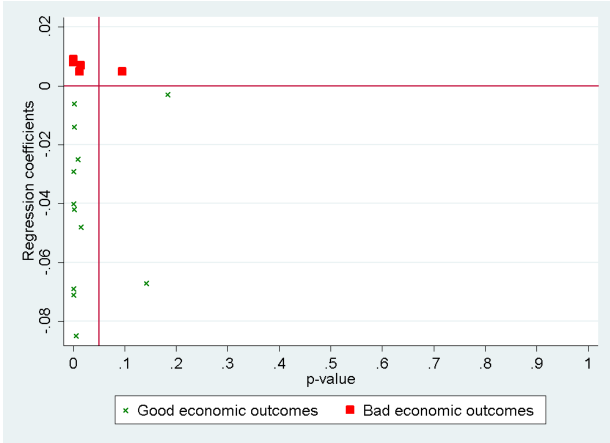
Table E6. Relationship between the Local ICS-30 and other socio-political outcomes

	(1)	(2)	(3)	(4)
Life expectancy				
Life expectancy at birth, males	-0.014	0.020	-0.060	0.182
	(0.176)	(0.171)	(0.173)	(0.121)
Life expectancy at 65, males	0.007	-0.005	-0.028	0.069
	(0.062)	(0.060)	(0.062)	(0.042)
Life expectancy at birth, females	-0.453	-0.292	-0.353	-0.080
	(0.205)**	(0.218)	(0.209)*	(0.164)
Life expectancy at 65, females	-0.202	-0.157	-0.169	-0.079
	(0.063)***	(0.072)**	(0.068)**	(0.055)
Crime Rates				
Total crimes	-0.005	0.001	0.001	-0.007
	(0.007)	(0.007)	(0.007)	(0.005)
Violent crimes	0.003	0.003	0.003	-0.002
	(0.007)	(0.007)	(0.007)	(0.005)
Thefts	-0.002	0.004	0.003	-0.002
	(0.005)	(0.005)	(0.005)	(0.003)
Other crimes	-0.006	-0.007	-0.005	-0.013
	(0.007)	(0.007)	(0.007)	(0.005)**
Murders per 100,000 inhabitants	0.005	0.002	0.003	0.000
	(0.002)*	(0.002)	(0.002)	(0.002)
Petty thefts per 100,000 inhabitants	-0.002	0.001	0.000	-0.002
	(0.002)	(0.002)	(0.002)	(0.001)
Snatching per 100,000 inhabitants	0.004	0.004	0.004	0.002
	(0.002)**	(0.002)**	(0.002)**	(0.001)
Burglaries per 100,000 inhabitants	-0.015	-0.012	-0.014	-0.009
	(0.004)***	(0.005)**	(0.006)**	(0.004)**
Theft of parked cars per 100,000 inhabitants	-0.006	0.000	-0.001	-0.002
	(0.004)	(0.004)	(0.004)	(0.003)
Car thefts per 100,000 inhabitants	0.006	0.004	0.004	0.000
	(0.002)**	(0.002)*	(0.002)*	(0.002)
Scams per 100,000 inhabitants	-0.010	-0.007	-0.008	-0.007
	(0.005)*	(0.005)	(0.005)	(0.004)**
Smuggling offences per 100,000 inhabitants	0.003	0.003	0.003	0.001
	(0.001)***	(0.001)***	(0.001)***	(0.001)
Drug production and sale per 100,000 inhabitants	-0.001	0.000	0.000	-0.002
	(0.004)	(0.004)	(0.004)	(0.003)
Exploitation of prostitution per 100,000 inhabitants	-0.010	-0.009	-0.009	-0.006
	(0.003)***	(0.003)***	(0.003)***	(0.002)***
Distraints per 1,000 inhabitants aged 18+	0.000	-0.001	-0.001	0.001
	(0.004)	(0.004)	(0.004)	(0.002)
Distraints per 1,000 families	0.002	0.000	-0.001	0.002
	(0.004)	(0.004)	(0.004)	(0.002)
Suicides Rates				
Suicides per 100,000 - Total	-0.014	-0.011	-0.012	-0.006
	(0.004)***	(0.005)**	(0.005)**	(0.004)
Suicides per 100,000 population - Males	-0.004	-0.002	-0.002	0.000
	(0.002)*	(0.002)	(0.002)	(0.002)
Suicides per 100,000 population - Females	-0.004	-0.002	-0.002	-0.001
	(0.003)	(0.003)	(0.003)	(0.002)
Suicide attempts per 100,000 - Total	-0.006	-0.002	-0.002	0.000
	(0.003)*	(0.004)	(0.004)	(0.003)
Suicide attempts per 100,000 - Males	-0.003	-0.001	0.000	-0.001
	(0.002)	(0.003)	(0.003)	(0.002)
Suicide attempts per 100,000 - Females	-0.002	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.001)
Public sector activity				
Value of public works started (pct VA)	0.005	0.004	0.005	0.004
	(0.002)*	(0.002)	(0.002)**	(0.002)***
Value of public works started by provinces (pct VA)	0.005	0.005	0.005	0.003
	(0.002)***	(0.002)***	(0.002)***	(0.001)***
Value of public works started (construction sector, pct VA)	0.010	0.009	0.010	0.006
	(0.003)***	(0.003)***	(0.002)***	(0.002)***
Value of public works completed (pct VA)	0.005	0.002	0.004	0.002
	(0.003)	(0.004)	(0.003)	(0.002)
Value of public works completed by provinces (pct VA)	0.007	0.006	0.006	0.004
	(0.002)***	(0.002)***	(0.002)***	(0.001)***
Percentage politicians with at least secondary education	0.013	0.004	0.007	0.017
	(0.017)	(0.017)	(0.017)	(0.012)
Ratio of paid to committed expenses	0.020	0.033	0.035	-0.003
	(0.055)	(0.054)	(0.054)	(0.038)
Deficit per capita in euros	0.003	0.004	0.004	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Growth rate of deficit per capita in euros	0.001	0.001	0.001	0.000
	(0.001)	(0.002)	(0.001)	(0.001)
Controls:				
Value added per capita	NO	YES	NO	NO
North/South dummy	NO	NO	YES	NO
Net migration flows (avg. 1999-2002)	NO	NO	NO	YES

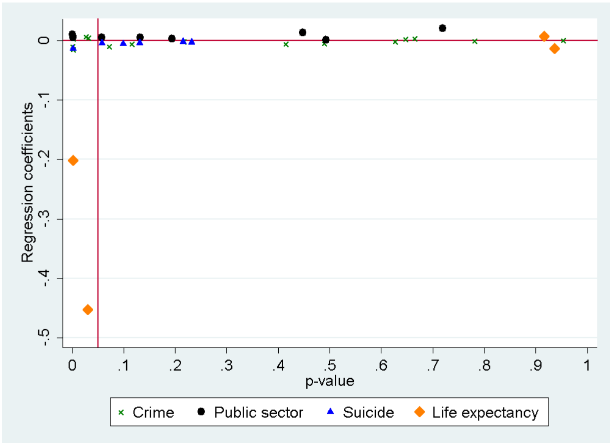
Notes: Each coefficient is obtained from a separate regression of the Local ICS-30 on each variable. Local ICS-30 refers to the ICS calculated including only surnames that contain at most 30 people (see Section 5) and are local (see Section 2.2). Standard errors in parentheses. (***) indicates significance at the 1% level, (**) indicates significance at the 5% level and (*) indicates significance at the 10% level.



(a) Key outcomes



(b) Other economic outcomes



(c) Other socio-political outcomes

FIGURE E2. Coefficients and p-values from separate regressions of the Local ICS-30 on province-level outcomes

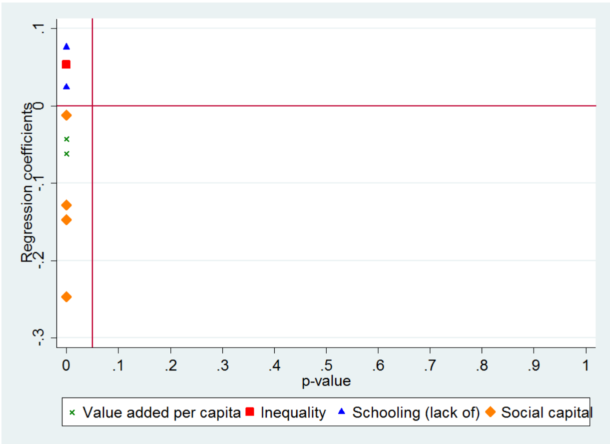
F Appendix: Robustness checks using alternative populations

This appendix provides results using the same ICS as in the body of the paper, namely the ICS-30 calculated with those individuals whose surname contains no more than 30 people, but making different choices as to the selection of the population. Specifically, we include both males and females and, alternatively, exclude the self-employed. Our results carry over to these alternative populations. Table F1 shows that the correlation between the ICS calculated on the different populations is very high. More specifically, Figure F1 shows the unconditional results including females and Figure F2 the unconditional results excluding self-employed workers who are more likely to under-report. Results are very close to those presented in Section 6. Results controlling for value added per capita, a North/South dummy and net migration flows are also similar to those presented in the body of the paper and are available upon request.

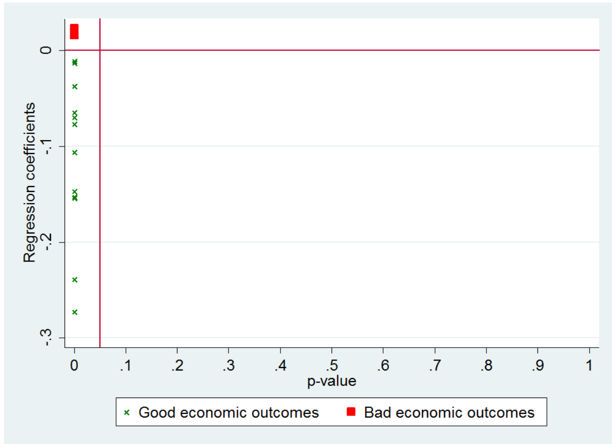
Table F1. Pairwise correlations across ICS measures

	ICS-30 males	ICS-30 males and females	ICS-30 no self-employed
ICS-30 males	1.0000		
ICS-30 males and females	0.9174	1.0000	
ICS-30 no self-employed	0.8251	0.8019	1.0000

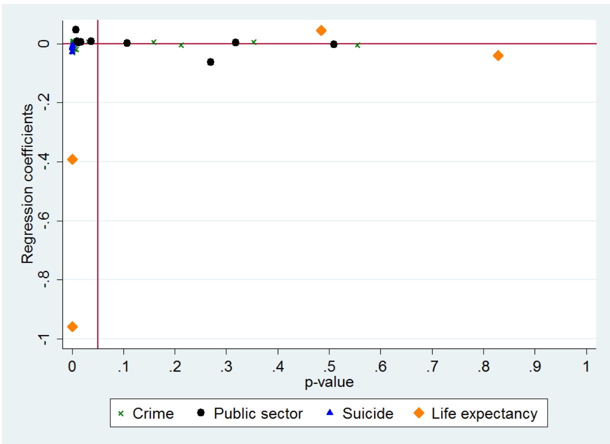
Notes: ICS-30 refers to the ICS calculated including only surnames that contain at most 30 people. Source: 2005 Italian tax records.



(a) Key outcomes

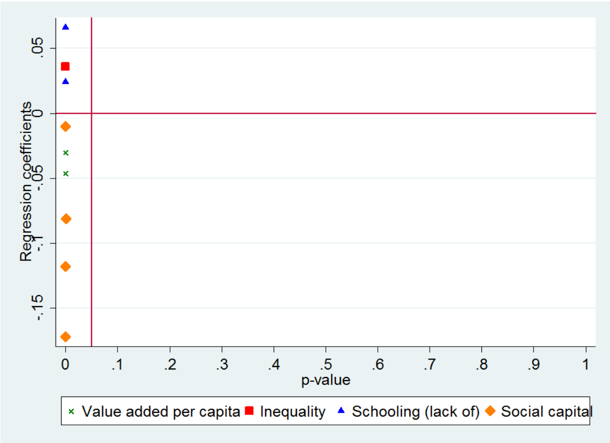


(b) Other economic outcomes

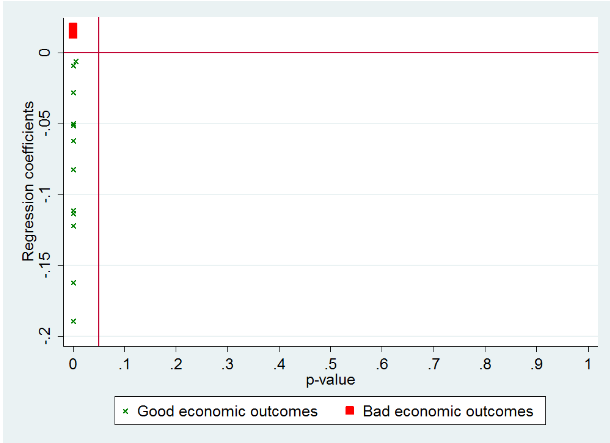


(c) Other socio-political outcomes

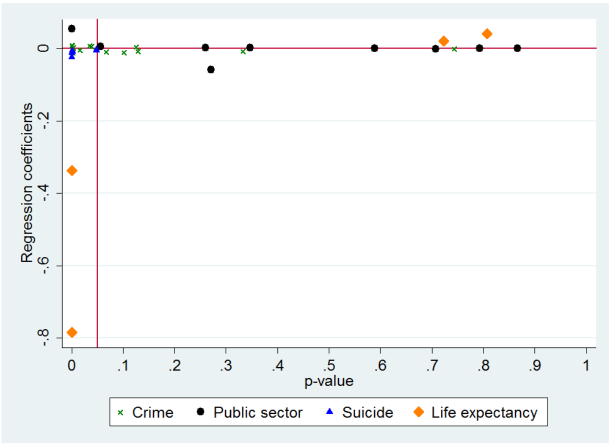
FIGURE F1. Coefficients and p-values from separate regressions of the ICS-30 computed including females on province-level outcomes



(a) Key outcomes



(b) Other economic outcomes



(c) Other socio-political outcomes

FIGURE F2. Coefficients and p-values from separate regressions of the ICS-30 computed excluding self-employed workers on province-level outcomes